

Package: paweh (via r-universe)

September 17, 2026

Title Power and Sample Size Analysis for Genetic Studies with Heterogeneity and Misclassification

Version 0.99.0

Description Provides prospective power and minimum sample size necessary (MSSN) calculations for genetic and genomic association study design. Supported designs include case-control studies, transmission disequilibrium tests with affected-child trios, and single- and multivariate quantitative-trait studies. The package emphasizes selected departures from idealized assumptions, including locus heterogeneity and supported forms of genotype and phenotype misclassification. These factors can materially affect power and sample-size requirements but can be difficult to incorporate analytically. Published statistical-genetics methods are implemented in a reproducible, primarily parameter- and model-based framework with static visualizations and an optional Shiny dashboard for interactive exploration. The software supports prospective study planning; it does not perform downstream association testing.

License MIT + file LICENSE

Encoding UTF-8

URL <https://github.com/akilanthony/paweh>,
<https://akilanthony.github.io/paweh/>

BugReports <https://github.com/akilanthony/paweh/issues>

biocViews ExperimentalDesign, Genetics, GenomeWideAssociation, ShinyApps

Roxygen list(markdown = TRUE)

Suggests BiocStyle, knitr, plotly, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports bslib, ggplot2, mvtnorm, rlang, shiny, withr

VignetteBuilder knitr

Config/roxygen2/version 8.1.0
Config/pak/sysreqs cmake make libuv1-dev zlib1g-dev
Repository <https://biocstaging.r-universe.dev>
Date/Publication 2026-09-17 13:18:55 UTC
RemoteUrl <https://github.com/BiocStaging/paweh>
RemoteRef HEAD
RemoteSha 7a5ed7bdfcab9f9f63e7e260204058bfcadf9cce

Contents

case_control_genotype_misclassification	3
case_control_locus_heterogeneity	8
case_control_phenotype_misclassification	11
cc_mssn	13
cc_ngs_mssn	17
cc_ngs_power	21
cc_power	24
lrtae_mssn	29
lrtae_ncp	31
lrtae_power	33
ngs_genotype_error_matrix	35
paweh_app	36
plot_cc_mssn	37
plot_cc_power	38
plot_ngs_mssn	40
plot_ngs_power	42
plot_qtl_anova_mssn	43
plot_qtl_anova_power	44
plot_qtl_genotype_distribution	45
plot_qtl_threshold_chisq_mssn	47
plot_qtl_threshold_chisq_power	49
plot_tdt_mssn	50
plot_tdt_mssn_locus_heterogeneity	52
plot_tdt_mssn_phenotype_misclassification	53
plot_tdt_power	55
plot_tdt_power_locus_heterogeneity	56
plot_tdt_power_phenotype_misclassification	58
plot_tdt_surface3d	59
qtl_anova_mssn	62
qtl_anova_power	63
qtl_falconer_parameters	65
qtl_falconer_threshold_parameters	66
qtl_multivariate_mssn_full	67
qtl_multivariate_power_full	69
qtl_threshold_chisq_mssn	71
qtl_threshold_chisq_power	73

<i>case_control_genotype_misclassification</i>	3
--	---

tdt_expected_nontransmission_probability	74
tdt_expected_transmission_counts	76
tdt_expected_transmission_probability	78
tdt_mssn	80
tdt_mssn_from_model	83
tdt_mssn_phenotype_misclassification_from_ncp	85
tdt_ngs_mssn	87
tdt_ngs_power	90
tdt_power	93
tdt_power_from_expected_counts	97
tdt_power_from_model	98

Index	101
--------------	------------

case_control_genotype_misclassification	<i>Specialized Case-Control Genotype Misclassification Functions</i>
---	--

Description

Convenience functions for genotype-only case-control chi-square calculations under conditional model-based genotype frequencies and genotype misclassification. These functions are narrower than `cc_mssn` and `cc_power`: they compute only the genotype chi-square test for one specific misclassification model.

Usage

```
cc_chisq_mssn_genotype_misclassification_1p(
  power,
  alpha,
  prev,
  pd,
  R2,
  MOI = c("M", "D", "Rec"),
  k = 1,
  e = 0,
  verbose = TRUE
)

cc_chisq_power_genotype_misclassification_1p(
  N_case,
  alpha,
  prev,
  pd,
  R2,
  MOI = c("M", "D", "Rec"),
  k = 1,
  e = 0,
```

```
    verbose = TRUE
  )

cc_chisq_mssn_genotype_misclassification_2p(
  power,
  alpha,
  prev,
  pd,
  R2,
  MOI = c("M", "D", "Rec"),
  k = 1,
  e1 = 0,
  e2 = 0,
  verbose = TRUE
)

cc_chisq_power_genotype_misclassification_2p(
  N_case,
  alpha,
  prev,
  pd,
  R2,
  MOI = c("M", "D", "Rec"),
  k = 1,
  e1 = 0,
  e2 = 0,
  verbose = TRUE
)

cc_chisq_mssn_genotype_misclassification_3p(
  power,
  alpha,
  prev,
  pd,
  R2,
  MOI = c("M", "D", "Rec"),
  k = 1,
  e01 = 0,
  e02 = 0,
  e03 = 0,
  verbose = TRUE
)

cc_chisq_power_genotype_misclassification_3p(
  N_case,
  alpha,
  prev,
  pd,
```

```
R2,  
MOI = c("M", "D", "Rec"),  
k = 1,  
e01 = 0,  
e02 = 0,  
e03 = 0,  
verbose = TRUE  
)  
  
cc_chisq_mssn_genotype_misclassification_differential_3p(  
  power,  
  alpha,  
  prev,  
  pd,  
  R2,  
  MOI = c("M", "D", "Rec"),  
  k = 1,  
  case_e01 = 0,  
  case_e02 = 0,  
  case_e03 = 0,  
  ctrl_e01 = 0,  
  ctrl_e02 = 0,  
  ctrl_e03 = 0,  
  verbose = TRUE  
)  
  
cc_chisq_power_genotype_misclassification_differential_3p(  
  N_case,  
  alpha,  
  prev,  
  pd,  
  R2,  
  MOI = c("M", "D", "Rec"),  
  k = 1,  
  case_e01 = 0,  
  case_e02 = 0,  
  case_e03 = 0,  
  ctrl_e01 = 0,  
  ctrl_e02 = 0,  
  ctrl_e03 = 0,  
  verbose = TRUE  
)
```

Arguments

power	Numeric in (0, 1). Desired target power for MSSN functions.
alpha	Numeric in (0, 1). Significance level.
prev	Numeric in (0, 1). Disease prevalence.

pd	Numeric in (0, 1). Disease-allele frequency.
R2	Numeric > 0. Homozygote relative risk.
MOI	Character. Mode of inheritance: "M", "D", or "Rec".
k	Numeric > 0. Control-to-case ratio N_{ctrl}/N_{case} .
e	Numeric in [0, 0.5]. One-parameter symmetric error probability assigned to each adjacent off-diagonal call. The matrix diagonal is $1 - 2e$; textbook Eq. 2.5 uses a total error parameter $\epsilon = 2e$.
verbose	Logical. If TRUE, prints a formatted summary.
N_case	Numeric > 0. Number of cases for power functions.
e1, e2	Numeric. Two-parameter genotype misclassification rates.
e01, e02, e03	Numeric. Three-parameter non-differential genotype misclassification rates.
case_e01, case_e02, case_e03	Numeric. Case-specific three-parameter genotype misclassification rates for differential misclassification.
ctrl_e01, ctrl_e02, ctrl_e03	Numeric. Control-specific three-parameter genotype misclassification rates for differential misclassification.

Details

These are specialized convenience functions retained for users who want a focused genotype-only calculation. The all-in-one case-control functions are recommended when combining model-free inputs, locus heterogeneity, trend tests, or multiple modifier choices in one call.

Conditional genotype probabilities are constructed from prevalence, allele frequency, and relative risks using Chapter 1, Section 1.4.2 and Eqs. 1.6–1.7 (p. 13) of Gordon, Finch, and Kim (2020). The genotype test uses the two-degree-of-freedom NCP in Eq. 1.22 (p. 26). The error matrices are the one-, two-, and three-parameter families in Eqs. 2.5, 2.6, and 2.7 (pp. 57–58). Douglas et al. (2002) and Sobel et al. (2002) support the corresponding error-model families; Gordon et al. (2002) supports their use in case-control power and sample-size calculations.

For the 1p functions, package e is the probability in each adjacent off-diagonal cell, and the diagonal is $1 - 2e$. Textbook Eq. 2.5 instead writes each off-diagonal as $\epsilon/2$ and the diagonal as $1 - \epsilon$; therefore $\epsilon = 2e$.

Differential 3p functions use separate Eq. 2.7-family matrices for cases and controls, as discussed in Chapter 2, Section 2.5.2 (pp. 61–69). Their power is nominal asymptotic power, and their MSSN is nominal MSSN, evaluated with the usual chi-square critical value. Different case and control error mechanisms can distort the null distribution and inflate type I error. These functions do not independently recalibrate the null distribution for arbitrary differential genotyping error.

The returned objects keep internal effect-size components such as S and, for differential misclassification, per-genotype components, for validation and debugging.

Value

A list with class matching the function name. MSSN functions include target non-centrality parameter, case/control sample sizes, model parameters, penetrances, misclassification matrices, internal S values, and true and observed genotype frequencies. Power functions include case/control sample sizes, non-centrality parameter, power, model parameters, penetrances, misclassification matrices, internal S values, and true and observed genotype frequencies.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.
- Gordon, D., Finch, S. J., Nothnagel, M., & Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Human Heredity*, 54(1), 22–33. doi:10.1159/000066696.
- Douglas, J. A., Skol, A. D., & Boehnke, M. (2002). Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *American Journal of Human Genetics*, 70(2), 487–495. doi:10.1086/338919.
- Sobel, E., Papp, J. C., & Lange, K. (2002). Detection and integration of genotyping errors in statistical genetics. *American Journal of Human Genetics*, 70(2), 496–508. doi:10.1086/338920.
- Ahn, K., Gordon, D., & Finch, S. J. (2009). Increase of rejection rate in case-control studies with the differential genotyping error rates. *Statistical Applications in Genetics and Molecular Biology*, 8(1), Article 25. doi:10.2202/15446115.1429.

See Also

[cc_power](#), [cc_mssn](#), [case_control_locus_heterogeneity](#), and [case_control_phenotype_misclassification](#).

Examples

```
cc_chisq_mssn_genotype_misclassification_1p(
  power = 0.8, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D", e = 0.02, verbose = FALSE
)

cc_chisq_power_genotype_misclassification_1p(
  N_case = 500, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D", e = 0.02, verbose = FALSE
)

cc_chisq_mssn_genotype_misclassification_2p(
  power = 0.8, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D", e1 = 0.02, e2 = 0.01, verbose = FALSE
)

cc_chisq_power_genotype_misclassification_2p(
  N_case = 500, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D", e1 = 0.02, e2 = 0.01, verbose = FALSE
)

cc_chisq_mssn_genotype_misclassification_3p(
  power = 0.8, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D", e01 = 0.02, e02 = 0.01, e03 = 0.005,
  verbose = FALSE
)

cc_chisq_power_genotype_misclassification_3p(
  N_case = 500, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
```

```

    MOI = "D", e01 = 0.02, e02 = 0.01, e03 = 0.005,
    verbose = FALSE
)

cc_chisq_mssn_genotype_misclassification_differential_3p(
  power = 0.8, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D",
  case_e01 = 0.02, case_e02 = 0.01, case_e03 = 0.005,
  ctrl_e01 = 0.01, ctrl_e02 = 0.005, ctrl_e03 = 0.002,
  verbose = FALSE
)

cc_chisq_power_genotype_misclassification_differential_3p(
  N_case = 500, alpha = 0.05, prev = 0.1, pd = 0.3, R2 = 1.8,
  MOI = "D",
  case_e01 = 0.02, case_e02 = 0.01, case_e03 = 0.005,
  ctrl_e01 = 0.01, ctrl_e02 = 0.005, ctrl_e03 = 0.002,
  verbose = FALSE
)

```

case_control_locus_heterogeneity

Specialized Case-Control Locus-Heterogeneity Functions

Description

Convenience functions for case-control association calculations under locus heterogeneity. These focused functions compute genotype chi-square or genotype trend calculations one test at a time. The all-in-one case-control functions remain the recommended interface when multiple tests or modifiers are needed in one call.

Usage

```

cc_chisq_mssn_locus_heterogeneity(
  power,
  alpha,
  g_case_assoc,
  g_ctrl,
  pi,
  k = 1,
  verbose = TRUE
)

cc_chisq_power_locus_heterogeneity(
  N_case,
  alpha,
  g_case_assoc,

```

```

    g_ctrl,
    pi,
    k = 1,
    verbose = TRUE
)

cc_trend_mssn_locus_heterogeneity(
  power,
  alpha,
  g_case_assoc,
  g_ctrl,
  pi,
  k = 1,
  w = c(0, 1, 2),
  verbose = TRUE
)

cc_trend_power_locus_heterogeneity(
  N_case,
  alpha,
  g_case_assoc,
  g_ctrl,
  pi,
  k = 1,
  w = c(0, 1, 2),
  verbose = TRUE
)

```

Arguments

power	Numeric in (0, 1). Desired target power for MSSN functions.
alpha	Numeric in (0, 1). Significance level.
g_case_assoc	Numeric vector of length 3. Case genotype frequencies under the associated-locus model, ordered as $c(g_0, g_1, g_2)$ and summing to 1.
g_ctrl	Numeric vector of length 3. Control genotype frequencies, ordered as $c(g_0, g_1, g_2)$ and summing to 1.
pi	Numeric in $[0, 1]$. Locus-homogeneity fraction. The adjusted case genotype frequencies are $\pi g_{case,assoc} + (1 - \pi)g_{ctrl}$.
k	Numeric > 0 . Control-to-case ratio N_{ctrl}/N_{case} .
verbose	Logical. If TRUE, prints a formatted summary.
N_case	Numeric > 0 . Number of cases for power functions.
w	Numeric vector of length 3. Trend-test genotype scores. The three weights cannot all be equal.

Details

The associated case distribution is mixed with the control distribution as $\pi g_{case,assoc} + (1 - \pi)g_{ctrl}$, following textbook Eq. 2.16 (p. 88). Thus pi is the homogeneous/associated fraction and $1 - \pi$ is

the heterogeneous fraction. The adjusted probabilities are then passed to either the genotype chi-square NCP in Eq. 1.22 (p. 26) or the trend-test NCP in Eq. 1.24 (p. 27). The trend construction under locus heterogeneity is given more specifically by Eqs. 5.29a–b (pp. 287–288).

Value

A list with class matching the function name. MSSN functions include target non-centrality parameter, sample sizes, internal S values, and adjusted frequencies. Power functions include sample sizes, non-centrality parameter, power, internal S values, and adjusted frequencies. Trend functions additionally return the trend numerator and denominator.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.
- Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3), 375–386. doi:10.2307/3001775.
- Slager, S. L., & Schaid, D. J. (2001). Case-control studies of genetic markers: power and sample size approximations for Armitage’s test for trend. *Human Heredity*, 52(3), 149–153. doi:10.1159/000053370.

See Also

[cc_power](#), [cc_mssn](#), and [case_control_genotype_misclassification](#).

Examples

```
cc_chisq_mssn_locus_heterogeneity(
  power = 0.8, alpha = 0.05,
  g_case_assoc = c(0.25, 0.50, 0.25),
  g_ctrl = c(0.36, 0.48, 0.16),
  pi = 0.8,
  verbose = FALSE
)

cc_chisq_power_locus_heterogeneity(
  N_case = 500, alpha = 0.05,
  g_case_assoc = c(0.25, 0.50, 0.25),
  g_ctrl = c(0.36, 0.48, 0.16), pi = 0.8, verbose = FALSE
)

cc_trend_mssn_locus_heterogeneity(
  power = 0.8, alpha = 0.05,
  g_case_assoc = c(0.25, 0.50, 0.25),
  g_ctrl = c(0.36, 0.48, 0.16), pi = 0.8, verbose = FALSE
)

cc_trend_power_locus_heterogeneity(
  N_case = 500, alpha = 0.05,
  g_case_assoc = c(0.25, 0.50, 0.25),
  g_ctrl = c(0.36, 0.48, 0.16),
```

```

    pi = 0.8,
    verbose = FALSE
)

```

case_control_phenotype_misclassification

Specialized Case-Control Phenotype Misclassification Functions

Description

Convenience functions for genotype chi-square case-control calculations in the presence of phenotype misclassification only. These focused functions are narrower than [cc_mssn](#) and [cc_power](#): they use true affected and true unaffected genotype frequencies supplied directly by the user and do not apply locus heterogeneity or genotype misclassification.

Usage

```

cc_chisq_power_phenotype_misclassification(
  N_case,
  alpha = 0.05,
  g_aff,
  g_unaff,
  prev,
  theta = 0,
  phi = 0,
  k = 1
)

```

```

cc_chisq_mssn_phenotype_misclassification(
  target_power = 0.8,
  alpha = 0.05,
  g_aff,
  g_unaff,
  prev,
  theta = 0,
  phi = 0,
  k = 1
)

```

Arguments

N_case	Numeric > 0. Number of cases for power calculations.
alpha	Numeric in (0, 1). Significance level.
g_aff	Numeric vector of true affected genotype frequencies, ordered consistently across affected and unaffected groups and summing to 1.

<code>g_unaff</code>	Numeric vector of true unaffected genotype frequencies, ordered consistently with <code>g_aff</code> and summing to 1.
<code>prev</code>	Numeric in (0, 1). Disease prevalence.
<code>theta</code>	Numeric in [0, 1). Probability that a truly affected individual is classified as a control, $\Pr(\text{affected} \rightarrow \text{control})$.
<code>phi</code>	Numeric in [0, 1). Probability that a truly unaffected individual is classified as a case, $\Pr(\text{unaffected} \rightarrow \text{case})$.
<code>k</code>	Numeric > 0. Control-to-case ratio N_{ctrl}/N_{case} .
<code>target_power</code>	Numeric in (0, 1). Desired target power for MSSN calculations.

Details

Phenotype misclassification is applied by mixing the true affected genotype distribution `g_aff` and true unaffected genotype distribution `g_unaff` using disease prevalence and the phenotype-error probabilities. Here $\theta = \Pr(\text{affected} \rightarrow \text{control})$ and $\phi = \Pr(\text{unaffected} \rightarrow \text{case})$. These specialized functions apply only phenotype misclassification; use the full case-control functions for combined phenotype misclassification, locus heterogeneity, genotype misclassification, or trend tests. The observed case distribution is a prevalence-weighted mixture of true affected and unaffected genotype probabilities conditional on being classified as a case; the control distribution is constructed analogously.

Value

`cc_chisq_power_phenotype_misclassification()` returns a list containing sample sizes, phenotype-error parameters, observed case/control genotype frequencies, internal S, non-centrality parameter `lambda`, and power. `cc_chisq_mssn_phenotype_misclassification()` returns a list containing target power, phenotype-error parameters, observed case/control genotype frequencies, internal S, target non-centrality parameter `lambda_star`, and case/control/total MSSN.

References

Edwards, B. J., Haynes, C., Levenstien, M. A., Finch, S. J., & Gordon, D. (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genetics*, 6, 18. doi:10.1186/14712156618. PMID: 15819990.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[cc_power](#), [cc_mssn](#), and [case_control_genotype_misclassification](#).

Examples

```
g_aff <- c((1 - 0.05)^2, 2 * 0.05 * (1 - 0.05), 0.05^2)
g_unaff <- c((1 - 0.15)^2, 2 * 0.15 * (1 - 0.15), 0.15^2)

cc_chisq_power_phenotype_misclassification(
  N_case = 250, alpha = 0.01,
```

```

    g_aff = g_aff, g_unaff = g_unaff,
    prev = 0.05, theta = 0, phi = 0.01
  )

cc_chisq_mssn_phenotype_misclassification(
  target_power = 0.8, alpha = 0.01,
  g_aff = g_aff, g_unaff = g_unaff,
  prev = 0.05, theta = 0, phi = 0.01
)

```

cc_mssn

Case-Control Minimum Sample Size for Conditional Genotype Frequencies

Description

Computes the minimum sample size necessary (MSSN) for case-control association tests using conditional genotype frequencies. The function supports model-based genotype frequencies, model-free genotype frequencies, optional locus heterogeneity, optional phenotype misclassification, optional genotype misclassification, genotype chi-square tests, and genotype trend tests.

Usage

```

cc_mssn(
  power,
  alpha,
  input_mode = c("model_based", "model_free"),
  prev = NULL,
  pd = NULL,
  R2 = NULL,
  MOI = c("M", "D", "Rec"),
  g1 = NULL,
  g0 = NULL,
  locus_het = FALSE,
  pi = 1,
  pheno_misclass = FALSE,
  theta = 0,
  phi = 0,
  k = 1,
  w = c(0, 1, 2),
  geno_misclass = c("none", "1p", "2p", "3p", "diff3p"),
  e = 0,
  e1 = 0,
  e2 = 0,
  e01 = 0,
  e02 = 0,
  e03 = 0,

```

```

case_e01 = 0,
case_e02 = 0,
case_e03 = 0,
ctrl_e01 = 0,
ctrl_e02 = 0,
ctrl_e03 = 0,
diff_source = c("explicit", "case", "ctrl"),
diff_multiplier = 1,
verbose = TRUE
)

```

Arguments

power	Numeric in (0, 1). Desired target power.
alpha	Numeric in (0, 1). Significance level.
input_mode	Character. One of "model_based" or "model_free". See Details.
prev	Numeric in (0, 1). Disease prevalence for input_mode = "model_based".
pd	Numeric in (0, 1). Disease-allele frequency for input_mode = "model_based".
R2	Numeric > 0. Homozygote relative risk for input_mode = "model_based".
MOI	Character. Mode of inheritance for model-based frequencies: "M" for multiplicative, "D" for dominant, or "Rec" for recessive.
g1, g0	Numeric vectors of length 3 for input_mode = "model_free". g1 gives case genotype frequencies and g0 gives control genotype frequencies, each ordered as c(g0, g1, g2) and summing to 1.
locus_het	Logical. If TRUE, requests a separate locus- heterogeneity sensitivity scenario starting from the baseline frequencies.
pi	Numeric in [0, 1]. Locus-homogeneity fraction used when locus_het = TRUE. When locus_het = FALSE, pi must remain at its default value of 1.
pheno_misclass	Logical. If TRUE, requests a separate phenotype-misclassification sensitivity scenario starting from baseline.
theta	Numeric in [0, 1). Probability that a truly affected individual is classified as a control.
phi	Numeric in [0, 1). Probability that a truly unaffected individual is classified as a case.
k	Numeric > 0. Control-to-case sample size ratio N_{ctrl}/N_{case} .
w	Numeric vector of length 3. Genotype trend-test scores. The three weights cannot all be equal.
geno_misclass	Character. Genotype misclassification model: "none", "1p", "2p", "3p", or "diff3p".
e	Numeric in [0, 0.5]. For geno_misclass = "1p", the probability assigned to each adjacent off-diagonal genotype call; the diagonal probability is $1 - 2e$. Thus the ϵ in textbook Eq. 2.5 equals $2e$; see Details.
e1, e2	Numeric. Two-parameter misclassification rates for geno_misclass = "2p".

e01, e02, e03	Numeric. Non-differential three-parameter misclassification rates for geno_misclass = "3p".
case_e01, case_e02, case_e03	Numeric. Case-specific three-parameter misclassification rates for geno_misclass = "diff3p".
ctrl_e01, ctrl_e02, ctrl_e03	Numeric. Control-specific three-parameter misclassification rates for geno_misclass = "diff3p".
diff_source	Character. For geno_misclass = "diff3p", one of "explicit", "case", or "ctrl". See Details.
diff_multiplier	Numeric ≥ 0 . Multiplier used when diff_source = "case" or diff_source = "ctrl".
verbose	Logical. If TRUE, prints a clean formatted summary.

Details

The no-error scenario is always calculated from the baseline conditional case and control genotype frequencies. Each requested modifier is evaluated independently from that same baseline:

- no error: baseline -> tests;
- phenotype misclassification: baseline -> phenotype misclassification -> tests;
- locus heterogeneity: baseline -> locus heterogeneity -> tests;
- genotype misclassification: baseline -> genotype misclassification -> tests.

Supplying several modifiers requests several independent sensitivity scenarios. Ordinary cc_mssn() does not combine modifiers. Specialized joint-error methods such as LRTae and LTTae are separate methods and are not implemented here.

With input_mode = "model_based", conditional case and control genotype frequencies are derived from prev, pd, R2, and MOI using Chapter 1, Section 1.4.2, Eqs. 1.6–1.7 (p. 13) of Gordon, Finch, and Kim (2020). With input_mode = "model_free", the user supplies g1 and g0 directly; when phenotype misclassification is enabled, g1 is treated as the true affected genotype distribution and g0 is treated as the true unaffected genotype distribution.

Phenotype misclassification requires prev in both input modes because disease prevalence is used to mix the baseline affected and unaffected genotype distributions into observed case and control distributions. Locus heterogeneity uses $g_{case,het} = \pi g_{case,base} + (1 - \pi)g_{ctrl,base}$ and leaves controls unchanged. Genotype misclassification matrices are likewise applied directly to baseline case and control frequencies.

The genotype misclassification models are: "none" for identity matrices, "1p" for one symmetric error rate, "2p" for adjacent homozygote/heterozygote and heterozygote error rates, "3p" for a non-differential three-parameter matrix, and "diff3p" for separate case and control three-parameter matrices. The one-, two-, and three-parameter matrices correspond to textbook Eqs. 2.5, 2.6, and 2.7 (pp. 57–58). For "1p", package e is the probability in each adjacent off-diagonal cell and the diagonal is $1 - 2e$; textbook Eq. 2.5 uses off-diagonal $\epsilon/2$, so $\epsilon = 2e$.

The genotype test uses Eq. 1.22 (p. 26). The trend test is defined by Eqs. 1.20–1.21 (p. 24) and uses the NCP in Eq. 1.24 (p. 27). Locus heterogeneity follows Eq. 2.16 (p. 88): π is the homogeneous

fraction and $1 - \pi$ is the heterogeneous fraction. Its trend-test construction corresponds to Eqs. 5.29a–b (pp. 287–288).

For `geno_misclass = "diff3p"`, `diff_source = "explicit"` uses the case and control error parameters exactly as supplied. With `diff_source = "case"`, control parameters are computed by multiplying the case parameters by `diff_multiplier`. With `diff_source = "ctrl"`, case parameters are computed by multiplying the control parameters by `diff_multiplier`.

With `geno_misclass = "diff3p"`, returned MSSNs are nominal, asymptotic values evaluated with the usual chi-square critical value. Different case and control error mechanisms can distort the null distribution and inflate type I error. This function does not independently recalibrate that null distribution for arbitrary differential genotyping error.

Internal effect-size components S are retained in the returned object for validation and debugging, but are not printed in the clean verbose output.

Value

An object of class `"cc_mssn"`, containing:

alpha, target_power Requested significance level and power.

input_mode, k, w, locus_het Input mode, control-to-case ratio, trend scores, and locus-heterogeneity settings.

errors Genotype-error model and matrices, plus phenotype-error settings and intermediate frequencies.

model_info Model-based penetrances and risk-model inputs, or model-free identifying information.

scenarios Named independent results. `no_error` always exists; requested modifiers add `phenotype_misclassification`, `heterogeneity`, and `genotype_misclassification`. Each contains scenario frequencies, modifier metadata, and genotype/trend MSSN results.

tests, freqs, compatibility_scenario Compatibility fields. With zero or one modifier, they mirror the sole design previously returned. With several modifiers, they explicitly mirror `no_error`; use `scenarios` for each sensitivity result.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.
- Gordon, D., Finch, S. J., Nothnagel, M., & Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Human Heredity*, 54(1), 22–33. doi:10.1159/000066696.
- Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3), 375–386. doi:10.2307/3001775.
- Slager, S. L., & Schaid, D. J. (2001). Case-control studies of genetic markers: power and sample size approximations for Armitage’s test for trend. *Human Heredity*, 52(3), 149–153. doi:10.1159/000053370.
- Edwards, B. J., Haynes, C., Levenstien, M. A., Finch, S. J., & Gordon, D. (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genetics*, 6, 18. doi:10.1186/14712156618.

See Also

[cc_power](#), [case_control_genotype_misclassification](#), [case_control_locus_heterogeneity](#), and [case_control_phenotype_misclassification](#).

Examples

```
cc_mssn(
  power = 0.8, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  geno_misclass = "none",
  verbose = FALSE
)

cc_mssn(
  power = 0.8, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  locus_het = TRUE, pi = 0.8,
  verbose = FALSE
)

cc_mssn(
  power = 0.8, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  locus_het = TRUE, pi = 0.8,
  geno_misclass = "3p", e01 = 0.02, e02 = 0.01, e03 = 0.005,
  verbose = FALSE
)

cc_mssn(
  power = 0.8, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  prev = 0.05,
  pheno_misclass = TRUE, theta = 0.05, phi = 0.01,
  verbose = FALSE
)
```

Description

Computes the minimum sample size necessary (MSSN) for a model-based case-control sequencing trend design. It evaluates the same independent sequencing-only, phenotype, heterogeneity, and post-call genotype-error scenarios as [cc_ngs_power](#).

Usage

```
cc_ngs_mssn(
  power,
  alpha,
  prev,
  pd,
  R2,
  coverage = NULL,
  seq_error = NULL,
  MOI = c("M", "D", "Rec"),
  k = 1,
  verbose = TRUE,
  locus_het = FALSE,
  pi = 1,
  case_coverage = coverage,
  ctrl_coverage = coverage,
  case_seq_error = seq_error,
  ctrl_seq_error = seq_error,
  pheno_misclass = FALSE,
  theta = 0,
  phi = 0,
  geno_misclass = c("none", "1p", "2p", "3p", "diff3p"),
  e = 0,
  e1 = 0,
  e2 = 0,
  e01 = 0,
  e02 = 0,
  e03 = 0,
  case_e01 = 0,
  case_e02 = 0,
  case_e03 = 0,
  ctrl_e01 = 0,
  ctrl_e02 = 0,
  ctrl_e03 = 0,
  diff_source = c("explicit", "case", "ctrl"),
  diff_multiplier = 1
)
```

Arguments

power	Numeric in (0, 1). Requested power.
alpha	Numeric in (0, 1). Significance level.

prev	Numeric in (0, 1). Disease prevalence.
pd	Numeric in (0, 1). Disease-allele frequency.
R2	Numeric > 0. Homozygote relative risk.
coverage	Common positive integer fixed sequencing depth. May be omitted when both group-specific depths are supplied.
seq_error	Symmetric per-read sequencing-error probability in [0, 0.5). Common shorthand; may be omitted when both group-specific errors are supplied.
MOI	Character mode of inheritance: "M" for multiplicative, "D" for dominant, or "Rec" for recessive.
k	Numeric > 0. Planned control-to-case sample-size ratio.
verbose	Logical. If TRUE, print a concise result summary.
locus_het	Logical. If TRUE, apply the canonical PAWEH case-control locus-heterogeneity mixture before sequencing observation.
pi	Numeric in [0, 1]. Locus-homogeneity fraction used when locus_het = TRUE. One retains the original associated-case distribution; zero makes the case distribution equal to controls. When locus_het = FALSE, pi must remain at its default value of 1.
case_coverage, ctrl_coverage	Case/control fixed depths, respectively. Each defaults to coverage; an explicit value overrides it.
case_seq_error, ctrl_seq_error	Case/control per-read error probabilities. Each defaults to seq_error; an explicit value overrides it. Effective depths and errors obey the common parameter constraints.
pheno_misclass	Logical. If TRUE, apply the ordinary PAWEH case-control phenotype-misclassification model before sequencing.
theta	Numeric in [0, 1). Probability that a truly affected individual is classified as a control.
phi	Numeric in [0, 1). Probability that a truly unaffected individual is classified as a case.
geno_misclass	Character genotype-misclassification model: "none", "1p", "2p", "3p", or "diff3p". The selected model is applied after sequencing calls.
e	Numeric in [0, 0.5]. One-parameter symmetric error rate.
e1, e2	Numeric two-parameter homozygote-to-heterozygote and heterozygote-to-homozygote error rates.
e01, e02, e03	Numeric non-differential three-parameter error rates.
case_e01, case_e02, case_e03	Case-specific three-parameter error rates.
ctrl_e01, ctrl_e02, ctrl_e03	Control-specific three-parameter error rates.
diff_source	Character source for differential three-parameter errors: "explicit", "case", or "ctrl".
diff_multiplier	Nonnegative multiplier used to derive the other group's rates when diff_source is "case" or "ctrl".

Details

Locus heterogeneity uses $g_{case,H} = \pi g_{case} + (1 - \pi)g_{control}$ in its independent branch. The prevalence-weighted phenotype transformation from `cc_mssn` is applied directly to the common biological baseline in its branch. Case and control sequencing settings are then applied to the corresponding pre-sequencing groups. Conventional genotype misclassification is applied after sequencing only in its own branch. See `cc_ngs_power` for the scenario paths and matrix convention. When $\pi = 0$ and sequencing mechanisms are equal, no finite MSSN exists for target power greater than alpha if the observed contrast is zero. Differential mechanisms may create an observed contrast even under a biological null. MSSN is a nominal asymptotic calculation using the existing chi-square critical value, without separate null-distribution or Type I error recalibration.

The function numerically inverts the one-degree-of-freedom noncentral chi-square distribution only to obtain the target NCP. It then solves the Ahn/Chapman-Nam trend-test sample-size equation analytically. Planned cases are the ceiling of the continuous requirement and planned controls are `ceiling(k * MSSN_case)`, following the PAWEH convention. Achieved NCP and power are recomputed using these actual integer sample sizes, with a local boundary adjustment if required to ensure target attainment and integer minimality.

Scores are `c(0, 1, 2)` for "M", `c(0, 1, 1)` for "D", and `c(0, 0, 1)` for "Rec". Finite depth can cause genotype-call uncertainty even when `seq_error = 0`, because a true heterozygote can yield reads from only one allele.

This is an analytic study-design calculation, not LTTae,NGS, a raw-read latent-genotype likelihood or EM method, downstream association testing, or simulation.

Value

Invisibly, an object of class "cc_ngs_mssn" containing a scenario-first scenarios list. Each scenario records target NCP, D/Q, continuous and integer sample sizes, achieved NCP and power, rounding metadata, staged frequencies, and sequencing/error matrices. Legacy fields follow the `compatibility_scenario` convention documented for `cc_ngs_power`.

References

Ahn, K., Haynes, C., Kim, W., St. Fleur, R., Gordon, D., & Finch, S. J. (2007). The effects of SNP genotyping errors on the power of the Cochran-Armitage linear trend test for case/control association studies. *Annals of Human Genetics*, 71, 249–261. doi:10.1111/j.14691809.2006.00318.x.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics*. Springer. doi:10.1007/9783030611217.

Examples

```
cc_ngs_mssn(
  power = 0.80, alpha = 0.05,
  prev = 0.05, pd = 0.30, R2 = 1.8,
  coverage = 20, seq_error = 0.01,
  MOI = "M", verbose = FALSE
)
```

cc_ngs_power

*Analytic Power for a Case-Control Sequencing Study***Description**

Computes prospective asymptotic power for a model-based case-control sequencing design. Sequencing is the common reference mechanism. Optional phenotype misclassification, locus heterogeneity, and conventional genotype misclassification are evaluated as independent sensitivity scenarios from the same biological baseline and sequencing design.

Usage

```
cc_ngs_power(
  N_case,
  alpha,
  prev,
  pd,
  R2,
  coverage = NULL,
  seq_error = NULL,
  MOI = c("M", "D", "Rec"),
  k = 1,
  verbose = TRUE,
  locus_het = FALSE,
  pi = 1,
  case_coverage = coverage,
  ctrl_coverage = coverage,
  case_seq_error = seq_error,
  ctrl_seq_error = seq_error,
  pheno_misclass = FALSE,
  theta = 0,
  phi = 0,
  geno_misclass = c("none", "1p", "2p", "3p", "diff3p"),
  e = 0,
  e1 = 0,
  e2 = 0,
  e01 = 0,
  e02 = 0,
  e03 = 0,
  case_e01 = 0,
  case_e02 = 0,
  case_e03 = 0,
  ctrl_e01 = 0,
  ctrl_e02 = 0,
  ctrl_e03 = 0,
  diff_source = c("explicit", "case", "ctrl"),
  diff_multiplier = 1
```

)

Arguments

N_case	Numeric > 0. Number of cases.
alpha	Numeric in (0, 1). Significance level.
prev	Numeric in (0, 1). Disease prevalence.
pd	Numeric in (0, 1). Disease-allele frequency.
R2	Numeric > 0. Homozygote relative risk.
coverage	Common positive integer fixed sequencing depth. May be omitted when both group-specific depths are supplied.
seq_error	Symmetric per-read sequencing-error probability in [0, 0.5). Common shorthand; may be omitted when both group-specific errors are supplied.
MOI	Character mode of inheritance: "M" for multiplicative, "D" for dominant, or "Rec" for recessive.
k	Numeric > 0. Control-to-case sample-size ratio N_{ctrl}/N_{case} .
verbose	Logical. If TRUE, print a concise result summary.
locus_het	Logical. If TRUE, apply the canonical PAWEH case-control locus-heterogeneity mixture before sequencing observation.
pi	Numeric in [0, 1]. Locus-homogeneity fraction used when locus_het = TRUE. One retains the original associated-case distribution; zero makes the case distribution equal to controls. When locus_het = FALSE, pi must remain at its default value of 1.
case_coverage, ctrl_coverage	Case/control fixed depths, respectively. Each defaults to coverage; an explicit value overrides it.
case_seq_error, ctrl_seq_error	Case/control per-read error probabilities. Each defaults to seq_error; an explicit value overrides it. Effective depths and errors obey the common parameter constraints.
pheno_misclass	Logical. If TRUE, apply the ordinary PAWEH case-control phenotype-misclassification model before sequencing.
theta	Numeric in [0, 1). Probability that a truly affected individual is classified as a control.
phi	Numeric in [0, 1). Probability that a truly unaffected individual is classified as a case.
geno_misclass	Character genotype-misclassification model: "none", "1p", "2p", "3p", or "diff3p". The selected model is applied after sequencing calls.
e	Numeric in [0, 0.5]. One-parameter symmetric error rate.
e1, e2	Numeric two-parameter homozygote-to-heterozygote and heterozygote-to-homozygote error rates.
e01, e02, e03	Numeric non-differential three-parameter error rates.

case_e01, case_e02, case_e03	Case-specific three-parameter error rates.
ctrl_e01, ctrl_e02, ctrl_e03	Control-specific three-parameter error rates.
diff_source	Character source for differential three-parameter errors: "explicit", "case", or "ctrl".
diff_multiplier	Nonnegative multiplier used to derive the other group's rates when diff_source is "case" or "ctrl".

Details

Trend scores are selected from MOI: "M" uses $c(0, 1, 2)$, "D" uses $c(0, 1, 1)$, and "Rec" uses $c(0, 0, 1)$. Power is the upper-tail probability beyond the central one-degree-of-freedom chi-square critical value under a noncentral chi-square distribution with the calculated NCP.

Locus heterogeneity uses the same parameterization as ordinary PAWEH case-control design:

$$g_{case,H} = \pi g_{case} + (1 - \pi) g_{control},$$

with the control distribution unchanged. The scenario paths are: sequencing only, baseline to sequencing to test; phenotype, baseline to phenotype misclassification to sequencing to test; heterogeneity, baseline to locus heterogeneity to sequencing to test; and genotype, baseline to sequencing to conventional genotype misclassification to test. Supplying multiple modifiers returns multiple independent scenarios; modifiers are not combined. The same effective case/control coverage and sequencing-error settings are used in every scenario. Phenotype misclassification uses the same prevalence-weighted transformation as `cc_power`, with $\theta = \Pr(\text{affected} \rightarrow \text{control})$ and $\phi = \Pr(\text{unaffected} \rightarrow \text{case})$. The resulting observed-case and observed-control distributions enter sequencing. Separate row-true, column-called matrices are then constructed using each observed group's effective depth and error, and applied as $t(M_{case}) \% \% g_{case}$ and $t(M_{ctrl}) \% \% g_{control}$. The ordinary PAWEH 1p, 2p, 3p, or differential 3p genotype-error matrices are applied only in their independent scenario, after sequencing calls, with the same row-current, column-observed convention. Only equal sequencing mechanisms permit commuting a shared biological mixture and sequencing. At $\pi = 0$, equal mechanisms with no phenotype-induced difference give zero NCP and power equal to α .

Differential sequencing or genotype-misclassification mechanisms can create observed case/control differences even under a biological null. Results remain nominal asymptotic calculations using the existing chi-square critical value; the null distribution and Type I error are not separately recalibrated here.

This is an analytic study-design calculation applied to sequencing-derived called genotypes. It is not a raw-read likelihood or EM analysis, performs no downstream association testing, and uses no simulation. Even when `seq_error = 0`, finite depth can cause call uncertainty because a true heterozygote can yield reads from only one allele.

Value

Invisibly, an object of class "cc_ngs_power". The scenarios list always contains `sequencing_only` and each explicitly requested modifier scenario. Every scenario records staged frequencies, sequencing and error matrices, Ahn components D and Q, NCP, and power. Legacy top-level result

fields mirror sequencing only with zero or multiple optional modifiers and the requested modifier with exactly one; `compatibility_scenario` identifies it. The frequency list distinguishes post-sequencing calls from final post-genotype-misclassification values. The sequencing list records the four effective group parameters and `case_transition_matrix/ctrl_transition_matrix`. Legacy `coverage/seq_error` retain supplied common inputs (or NULL); `transition_matrix` aliases the case matrix, shared when mechanisms are equal. Use the two explicit matrices for differential designs.

References

Ahn, K., Haynes, C., Kim, W., St. Fleur, R., Gordon, D., & Finch, S. J. (2007). The effects of SNP genotyping errors on the power of the Cochran-Armitage linear trend test for case/control association studies. *Annals of Human Genetics*, 71, 249–261. doi:[10.1111/j.14691809.2006.00318.x](https://doi.org/10.1111/j.14691809.2006.00318.x).

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics*. Springer. doi:[10.1007/9783030611217](https://doi.org/10.1007/9783030611217).

Examples

```
cc_ngs_power(
  N_case = 1000, alpha = 0.05,
  prev = 0.05, pd = 0.30, R2 = 1.8,
  coverage = 20, seq_error = 0.01,
  MOI = "M", verbose = FALSE
)
```

cc_power

Case-Control Power for Conditional Genotype Frequencies

Description

Computes power for fixed case-control sample sizes using conditional genotype frequencies. The function supports model-based genotype frequencies, model-free genotype frequencies, optional locus heterogeneity, optional phenotype misclassification, optional genotype misclassification, genotype chi-square tests and genotype trend tests.

Usage

```
cc_power(
  N_case,
  alpha,
  input_mode = c("model_based", "model_free"),
  prev = NULL,
  pd = NULL,
  R2 = NULL,
  MOI = c("M", "D", "Rec"),
  g1 = NULL,
  g0 = NULL,
```

```

locus_het = FALSE,
pi = 1,
pheno_misclass = FALSE,
theta = 0,
phi = 0,
k = 1,
w = c(0, 1, 2),
geno_misclass = c("none", "1p", "2p", "3p", "diff3p"),
e = 0,
e1 = 0,
e2 = 0,
e01 = 0,
e02 = 0,
e03 = 0,
case_e01 = 0,
case_e02 = 0,
case_e03 = 0,
ctrl_e01 = 0,
ctrl_e02 = 0,
ctrl_e03 = 0,
diff_source = c("explicit", "case", "ctrl"),
diff_multiplier = 1,
verbose = TRUE
)

```

Arguments

N_case	Numeric > 0. Number of cases.
alpha	Numeric in (0, 1). Significance level.
input_mode	Character. One of "model_based" or "model_free". See Details.
prev	Numeric in (0, 1). Disease prevalence for input_mode = "model_based".
pd	Numeric in (0, 1). Disease-allele frequency for input_mode = "model_based".
R2	Numeric > 0. Homozygote relative risk for input_mode = "model_based".
MOI	Character. Mode of inheritance for model-based frequencies: "M" for multiplicative, "D" for dominant, or "Rec" for recessive.
g1, g0	Numeric vectors of length 3 for input_mode = "model_free". g1 gives case genotype frequencies and g0 gives control genotype frequencies, each ordered as c(g0, g1, g2) and summing to 1.
locus_het	Logical. If TRUE, requests a separate locus- heterogeneity sensitivity scenario starting from the baseline frequencies.
pi	Numeric in [0, 1]. Locus-homogeneity fraction used when locus_het = TRUE. When locus_het = FALSE, pi must remain at its default value of 1.
pheno_misclass	Logical. If TRUE, requests a separate phenotype-misclassification sensitivity scenario starting from baseline.
theta	Numeric in [0, 1). Probability that a truly affected individual is classified as a control.

phi	Numeric in $[0, 1)$. Probability that a truly unaffected individual is classified as a case.
k	Numeric > 0 . Control-to-case sample size ratio N_{ctrl}/N_{case} .
w	Numeric vector of length 3. Genotype trend-test scores. The three weights cannot all be equal.
geno_misclass	Character. Genotype misclassification model: "none", "1p", "2p", "3p", or "diff3p".
e	Numeric in $[0, 0.5]$. For <code>geno_misclass = "1p"</code> , the probability assigned to each adjacent off-diagonal genotype call. The textbook Eq. 2.5 parameter satisfies $\epsilon = 2e$; see Details.
e1, e2	Numeric. Two-parameter misclassification rates for <code>geno_misclass = "2p"</code> .
e01, e02, e03	Numeric. Non-differential three-parameter misclassification rates for <code>geno_misclass = "3p"</code> .
case_e01, case_e02, case_e03	Numeric. Case-specific three-parameter misclassification rates for <code>geno_misclass = "diff3p"</code> .
ctrl_e01, ctrl_e02, ctrl_e03	Numeric. Control-specific three-parameter misclassification rates for <code>geno_misclass = "diff3p"</code> .
diff_source	Character. For <code>geno_misclass = "diff3p"</code> , one of "explicit", "case", or "ctrl". See Details.
diff_multiplier	Numeric ≥ 0 . Multiplier used when <code>diff_source = "case"</code> or <code>diff_source = "ctrl"</code> .
verbose	Logical. If TRUE, prints a clean formatted summary.

Details

The no-error scenario is always calculated from the baseline conditional case and control genotype frequencies. Each requested modifier is evaluated independently from that same baseline:

- no error: baseline -> tests;
- phenotype misclassification: baseline -> phenotype misclassification -> tests;
- locus heterogeneity: baseline -> locus heterogeneity -> tests;
- genotype misclassification: baseline -> genotype misclassification -> tests.

Supplying several modifiers requests several independent sensitivity scenarios. Ordinary `cc_power()` does not combine modifiers. Specialized joint-error methods such as LRTae and LTTae are separate methods and are not implemented here.

With `input_mode = "model_based"`, conditional case and control genotype frequencies are derived from `prev`, `pd`, `R2`, and `MOI` using Chapter 1, Section 1.4.2, Eqs. 1.6–1.7 (p. 13) of Gordon, Finch, and Kim (2020). With `input_mode = "model_free"`, the user supplies `g1` and `g0` directly; when phenotype misclassification is enabled, `g1` is treated as the true affected genotype distribution and `g0` is treated as the true unaffected genotype distribution.

Phenotype misclassification requires `prev` in both input modes because disease prevalence is used to mix the baseline affected and unaffected genotype distributions into observed case and control

distributions. Locus heterogeneity uses $g_{case,het} = \pi g_{case,base} + (1 - \pi)g_{ctrl,base}$ and leaves controls unchanged. Genotype misclassification matrices are likewise applied directly to baseline case and control frequencies.

The genotype misclassification models are: "none" for identity matrices, "1p" for one symmetric error rate, "2p" for adjacent homozygote/heterozygote and heterozygote error rates, "3p" for a non-differential three-parameter matrix, and "diff3p" for separate case and control three-parameter matrices. These matrices correspond to Eqs. 2.5–2.7 (pp. 57–58). For "1p", package e is the probability in each adjacent off-diagonal cell and the diagonal is $1 - 2e$; textbook Eq. 2.5 uses off-diagonal $\epsilon/2$, so $\epsilon = 2e$. The genotype and trend tests use Eqs. 1.22 and 1.24 (pp. 26–27), with the trend statistic defined by Eqs. 1.20–1.21 (p. 24). Locus heterogeneity follows Eq. 2.16 (p. 88), where π is the homogeneous fraction; its trend-test form is given in Eqs. 5.29a–b (pp. 287–288).

For `geno_misclass = "diff3p"`, `diff_source = "explicit"` uses the case and control error parameters exactly as supplied. With `diff_source = "case"`, control parameters are computed by multiplying the case parameters by `diff_multiplier`. With `diff_source = "ctrl"`, case parameters are computed by multiplying the control parameters by `diff_multiplier`.

With `geno_misclass = "diff3p"`, returned power is nominal asymptotic power evaluated with the usual chi-square critical value. Different case and control error mechanisms can distort the null distribution and inflate type I error. This function does not independently recalibrate the null distribution for arbitrary differential genotyping error.

Internal effect-size components S are retained in the returned object for validation and debugging, but are not printed in the clean verbose output.

Value

An object of class "cc_power", containing:

alpha, N_case, N_ctrl, N_total Significance level and numbers of case, control, and total individuals.

input_mode, k, w, locus_het Input mode, control-to-case ratio, trend scores, and locus-heterogeneity settings.

errors Genotype-error model and matrices, plus phenotype-error settings and intermediate frequencies.

model_info Model-based penetrances and risk-model inputs, or model-free identifying information.

scenarios Named independent results. `no_error` always exists; requested modifiers add `phenotype_misclassification`, `heterogeneity`, and `genotype_misclassification`. Each contains scenario frequencies, modifier metadata, and genotype/trend power results.

tests, freqs, compatibility_scenario Compatibility fields. With zero or one modifier, they mirror the sole design previously returned. With several modifiers, they explicitly mirror `no_error`; use scenarios for each sensitivity result.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

Gordon, D., Finch, S. J., Nothnagel, M., & Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Human Heredity*, 54(1), 22–33. doi:10.1159/000066696.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3), 375–386. doi:10.2307/3001775.

Slager, S. L., & Schaid, D. J. (2001). Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Human Heredity*, 52(3), 149–153. doi:10.1159/000053370.

Edwards, B. J., Haynes, C., Levenstien, M. A., Finch, S. J., & Gordon, D. (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genetics*, 6, 18. doi:10.1186/14712156618.

See Also

[cc_mssn](#), [case_control_genotype_misclassification](#), [case_control_locus_heterogeneity](#), and [case_control_phenotype_misclassification](#).

Examples

```
cc_power(
  N_case = 500, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  geno_misclass = "none",
  verbose = FALSE
)

cc_power(
  N_case = 500, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  locus_het = TRUE, pi = 0.8,
  verbose = FALSE
)

cc_power(
  N_case = 500, alpha = 0.05,
  input_mode = "model_free",
  g1 = c(0.25, 0.50, 0.25),
  g0 = c(0.36, 0.48, 0.16),
  locus_het = TRUE, pi = 0.8,
  geno_misclass = "3p", e01 = 0.02, e02 = 0.01, e03 = 0.005,
  verbose = FALSE
)

cc_power(
  N_case = 500, alpha = 0.05,
  input_mode = "model_based",
  prev = 0.05, pd = 0.30, R2 = 1.8, MOI = "M",
```

```

    pheno_misclass = TRUE, theta = 0.05, phi = 0.01,
    verbose = FALSE
  )

  cc_power(
    N_case = 500, alpha = 0.05,
    input_mode = "model_based",
    prev = 0.05, pd = 0.30, R2 = 1.8, MOI = "M",
    locus_het = TRUE, pi = 0.8,
    geno_misclass = "3p", e01 = 0.02, e02 = 0.01, e03 = 0.005,
    verbose = FALSE
  )

  cc_power(
    N_case = 500, alpha = 0.05,
    input_mode = "model_free",
    g1 = c(0.25, 0.50, 0.25),
    g0 = c(0.36, 0.48, 0.16),
    geno_misclass = "diff3p",
    diff_source = "explicit",
    case_e01 = 0.02, case_e02 = 0.01, case_e03 = 0.005,
    ctrl_e01 = 0.01, ctrl_e02 = 0.005, ctrl_e03 = 0.002,
    verbose = FALSE
  )

  cc_power(
    N_case = 500, alpha = 0.05,
    input_mode = "model_free",
    g1 = c(0.25, 0.50, 0.25),
    g0 = c(0.36, 0.48, 0.16),
    geno_misclass = "diff3p",
    diff_source = "case", diff_multiplier = 0.5,
    case_e01 = 0.02, case_e02 = 0.01, case_e03 = 0.005,
    verbose = FALSE
  )

```

Lrtae_mssn

Minimum Total Sample Size for an LRTae Design

Description

Finds the minimum integer total sample size required to attain a requested prospective LRTae power while preserving a requested allocation among the four measurement and validation tiers as closely as integer counts permit. This is a prospective-design calculation and is not an observed-data test.

Usage

```

Lrtae_mssn(
  true_control_genotype_frequencies,

```

```

true_case_genotype_frequencies,
true_sampling_proportions,
phenotype_misclassification_matrix,
genotype_misclassification_matrix,
target_power,
alpha,
prop_fallible,
prop_pheno_validated,
prop_geno_validated,
prop_both_validated
)

```

Arguments

<code>true_control_genotype_frequencies</code>	Numeric vector of true genotype frequencies among controls.
<code>true_case_genotype_frequencies</code>	Numeric vector of true genotype frequencies among cases, in the same category order as <code>true_control_genotype_frequencies</code> .
<code>true_sampling_proportions</code>	Numeric vector of length two containing the true control and case sampling proportions, in that order.
<code>phenotype_misclassification_matrix</code>	Numeric two-by-two matrix whose rows are true phenotype categories (control, case) and whose columns are observed phenotype categories (control, case).
<code>genotype_misclassification_matrix</code>	Numeric square matrix whose rows are true genotype categories and whose columns are observed genotype categories, in the genotype-frequency order.
<code>target_power</code>	Requested power, strictly greater than alpha and less than one.
<code>alpha</code>	Significance level, strictly between zero and one.
<code>prop_fallible</code>	Proportion allocated to fallible phenotype and genotype measurements only.
<code>prop_pheno_validated</code>	Proportion allocated to phenotype validation with fallible genotype measurement.
<code>prop_geno_validated</code>	Proportion allocated to genotype validation with fallible phenotype measurement.
<code>prop_both_validated</code>	Proportion allocated to validation of both phenotype and genotype.

Details

The four allocation proportions must be nonnegative and sum to one. For every candidate total N , integer counts are assigned by the largest-remainder (Hamilton) rule: floor $N * \text{proportion}$ for each tier, then distribute the remaining individuals in descending order of fractional remainder. Exact ties are resolved in the fixed order: fallible, phenotype validated, genotype validated, both validated. Thus allocated counts always sum exactly to N .

The search doubles an upper bound until it attains `target_power`, applies integer binary search within the bracket, and explicitly verifies the neighboring design at $N - 1$. Power at each candidate is evaluated through `lrtae_power()`, which uses the validated corrected `lrtae_ncp()` engine, $k - 1$ degrees of freedom, and a noncentral chi-square distribution.

Value

An object of class `lrtae_mssn`, containing `MSSN_total`, `target_power`, `achieved_power`, `achieved_ncp`, `df`, `alpha`, `critical_value`, requested allocation proportions, and the four allocated integer tier counts.

Examples

```
lrtae_mssn(
  true_control_genotype_frequencies = c(0.81, 0.18, 0.01),
  true_case_genotype_frequencies = c(0.64, 0.32, 0.04),
  true_sampling_proportions = c(0.5, 0.5),
  phenotype_misclassification_matrix = matrix(
    c(0.95, 0.05, 0.05, 0.95), 2, byrow = TRUE
  ),
  genotype_misclassification_matrix = matrix(
    c(0.950, 0.050, 0.000,
      0.025, 0.950, 0.025,
      0.000, 0.050, 0.950),
    3, byrow = TRUE
  ),
  target_power = 0.80,
  alpha = 0.05,
  prop_fallible = 0.7,
  prop_pheno_validated = 0.1,
  prop_geno_validated = 0.1,
  prop_both_validated = 0.1
)
```

Lrtae_ncp

Noncentrality Parameter for Prospective LRTae Designs

Description

Computes the analytic noncentrality parameter (NCP) for the likelihood-ratio test allowing for error (LRTae) in a prospective case-control design. The calculation incorporates fallible measurements and validation subsamples for phenotype, genotype, or both. It does not analyze observed study data.

Usage

```
lrtae_ncp(
  true_control_genotype_frequencies,
  true_case_genotype_frequencies,
```

```

    true_sampling_proportions,
    phenotype_misclassification_matrix,
    genotype_misclassification_matrix,
    n_fallible,
    n_pheno_validated,
    n_geno_validated,
    n_both_validated
)

```

Arguments

`true_control_genotype_frequencies`
 Numeric vector of true genotype frequencies among controls.

`true_case_genotype_frequencies`
 Numeric vector of true genotype frequencies among cases, in the same category order as `true_control_genotype_frequencies`.

`true_sampling_proportions`
 Numeric vector of length two containing the true control and case sampling proportions, in that order.

`phenotype_misclassification_matrix`
 Numeric two-by-two matrix whose rows are true phenotype categories (control, case) and whose columns are observed phenotype categories (control, case).

`genotype_misclassification_matrix`
 Numeric square matrix whose rows are true genotype categories and whose columns are observed genotype categories, in the genotype-frequency order.

`n_fallible`
 Number sampled with fallible phenotype and genotype measurements only.

`n_pheno_validated`
 Number whose phenotype is validated while genotype is measured fallibly.

`n_geno_validated`
 Number whose genotype is validated while phenotype is measured fallibly.

`n_both_validated`
 Number whose phenotype and genotype are both validated.

Value

A single nonnegative numeric value: the LRTae NCP.

References

- Ji, F., Yang, Y., Haynes, C., Finch, S. J., & Gordon, D. (2005). Computing asymptotic power and sample size for case-control genetic association studies in the presence of phenotype and/or genotype misclassification errors. *Statistical Applications in Genetics and Molecular Biology*, 4(1), Article 37. doi:[10.2202/15446115.1184](https://doi.org/10.2202/15446115.1184)
- Barral, S., Haynes, C., Stone, M., & Gordon, D. (2006). LRTae: improving statistical power for genetic association with case/control data when phenotype and/or genotype misclassification errors are present. *BMC Genetics*, 7, 24. doi:[10.1186/14712156724](https://doi.org/10.1186/14712156724)
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:[10.1007/9783030611217](https://doi.org/10.1007/9783030611217)

Examples

```

lrtae_ncp(
  true_control_genotype_frequencies = c(0.81, 0.18, 0.01),
  true_case_genotype_frequencies = c(0.64, 0.32, 0.04),
  true_sampling_proportions = c(0.5, 0.5),
  phenotype_misclassification_matrix = matrix(
    c(0.95, 0.05, 0.05, 0.95), nrow = 2, byrow = TRUE
  ),
  genotype_misclassification_matrix = matrix(
    c(0.950, 0.050, 0.000,
      0.025, 0.950, 0.025,
      0.000, 0.050, 0.950),
    nrow = 3, byrow = TRUE
  ),
  n_fallible = 700,
  n_pheno_validated = 100,
  n_geno_validated = 100,
  n_both_validated = 100
)

```

lrtae_power

Prospective Power for an LRTae Design

Description

Computes prospective power for the likelihood-ratio test allowing for error (LRTae). The function uses `lrtae_ncp()` and the corrected LRTae information matrix; it does not perform an observed-data hypothesis test.

Usage

```

lrtae_power(
  true_control_genotype_frequencies,
  true_case_genotype_frequencies,
  true_sampling_proportions,
  phenotype_misclassification_matrix,
  genotype_misclassification_matrix,
  n_fallible,
  n_pheno_validated,
  n_geno_validated,
  n_both_validated,
  alpha
)

```

Arguments

`true_control_genotype_frequencies`
 Numeric vector of true genotype frequencies among controls.

true_case_genotype_frequencies
 Numeric vector of true genotype frequencies among cases, in the same category order as true_control_genotype_frequencies.

true_sampling_proportions
 Numeric vector of length two containing the true control and case sampling proportions, in that order.

phenotype_misclassification_matrix
 Numeric two-by-two matrix whose rows are true phenotype categories (control, case) and whose columns are observed phenotype categories (control, case).

genotype_misclassification_matrix
 Numeric square matrix whose rows are true genotype categories and whose columns are observed genotype categories, in the genotype-frequency order.

n_fallible
 Number sampled with fallible phenotype and genotype measurements only.

n_pheno_validated
 Number whose phenotype is validated while genotype is measured fallibly.

n_geno_validated
 Number whose genotype is validated while phenotype is measured fallibly.

n_both_validated
 Number whose phenotype and genotype are both validated.

alpha
 Significance level, strictly between zero and one.

Details

For k genotype categories, the asymptotic reference distribution has $k - 1$ degrees of freedom. The rejection threshold is the upper α quantile of the central chi-square distribution, and power is the corresponding upper-tail probability from a noncentral chi-square distribution with the NCP returned by `lrtae_ncp()`.

Value

An object of class `lrtae_power`, containing `power`, `ncp`, `df`, `alpha`, `critical_value`, the four sampling-tier counts, and `N_total`.

Examples

```
lrtae_power(
  true_control_genotype_frequencies = c(0.81, 0.18, 0.01),
  true_case_genotype_frequencies = c(0.64, 0.32, 0.04),
  true_sampling_proportions = c(0.5, 0.5),
  phenotype_misclassification_matrix = matrix(
    c(0.95, 0.05, 0.05, 0.95), 2, byrow = TRUE
  ),
  genotype_misclassification_matrix = matrix(
    c(0.950, 0.050, 0.000,
      0.025, 0.950, 0.025,
      0.000, 0.050, 0.950),
    3, byrow = TRUE
  ),
  n_fallible = 700,
```

```

    n_pheno_validated = 100,
    n_geno_validated = 100,
    n_both_validated = 100,
    alpha = 0.05
)

```

ngs_genotype_error_matrix

Sequencing-Derived Genotype Error Matrix

Description

Constructs the analytic genotype transition matrix induced by fixed sequencing coverage, symmetric per-read sequencing error, and deterministic maximum-likelihood genotype calling.

Usage

```
ngs_genotype_error_matrix(coverage, seq_error)
```

Arguments

coverage	A single finite integer greater than or equal to 1 giving sequencing coverage (read depth).
seq_error	A single finite numeric sequencing-error probability in $[0, 0.5)$.

Details

For true genotype $G \in \{0, 1, 2\}$, the alternate-read count follows

$$X \mid G, V, \epsilon \sim \text{Binomial}(V, q_G),$$

where $q_0 = \epsilon$, $q_1 = 0.5$, and $q_2 = 1 - \epsilon$. For every possible alternate-read count, the called genotype maximizes its binomial likelihood among the three genotype models.

Numerical likelihood ties within $1e-12$ are resolved symmetrically: below an alternate-read fraction of one half, the lower tied genotype is selected; above one half, the higher tied genotype is selected; at exactly one half, the tied genotype closest to 1 is selected.

Rows correspond to true genotypes and columns to called genotypes. Thus, if E is the returned matrix and g_true is a length-three vector of true genotype probabilities, called-genotype probabilities are obtained as `as.numeric(t(E) %*% g_true)`. Every row sums to one within numerical precision.

At finite coverage, even with zero sequencing error, a true heterozygote can produce all-reference or all-alternate reads. Consequently the finite-depth matrix need not be exactly the identity, especially at very low coverage.

Value

A numeric 3×3 matrix. Rows `true_0`, `true_1`, and `true_2` identify the true genotype; columns `called_0`, `called_1`, and `called_2` identify the maximum-likelihood call.

Examples

```
E <- ngs_genotype_error_matrix(coverage = 10, seq_error = 0.01)
g_true <- c(0.70, 0.25, 0.05)
as.numeric(t(E) %*% g_true)
```

paweh_app

Construct the paweh Shiny Dashboard

Description

Creates the modular bslib dashboard for paweh study design. The Case-Control and TDT / Family workspaces provide canonical power, minimum-sample-size, sensitivity, and study-specific visual results. The Quantitative Trait workspace collects the canonical function inputs and displays that function's verbose console output. Dashboard calculations delegate to the same canonical package functions as the programmatic R interface.

Usage

```
paweh_app()
```

Details

paweh_app() returns the application object without launching it. This makes the app safe to construct in package code, tests, and deployment tooling. Launch it explicitly with [shiny::runApp\(\)](#).

Value

A Shiny application object inheriting from shiny.appobj.

Examples

```
app <- paweh_app()
inherits(app, "shiny.appobj")
if (interactive()) {
  shiny::runApp(app)
}
```

plot_cc_mssn

*Plot Case-Control Minimum Sample Size Necessary***Description**

Sweeps one x-axis parameter and repeatedly calls `cc_mssn()` to plot the minimum sample size necessary (MSSN) for case-control tests.

Usage

```
plot_cc_mssn(
  x_var,
  x_values,
  test = c("genotypes", "trend"),
  input_mode = c("model_based", "model_free"),
  sample_size = c("total", "case", "control"),
  compare_tests = FALSE,
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

<code>x_var</code>	Character. Parameter to vary on the x-axis.
<code>x_values</code>	Numeric vector of x-axis values.
<code>test</code>	One of "genotypes" or "trend".
<code>input_mode</code>	One of "model_based" or "model_free".
<code>sample_size</code>	One of "total", "case", or "control".
<code>compare_tests</code>	Logical. If TRUE, plot genotype and trend tests together and ignore test.
<code>title</code>	Optional character title override.
<code>x_label</code>	Optional x-axis label override.
<code>y_label</code>	Optional y-axis label override.
<code>return_data</code>	Logical. If TRUE, return the data frame instead of a ggplot.
<code>...</code>	Arguments passed to <code>cc_mssn()</code> .

Details

The supported test values are "genotypes" and "trend". `sample_size` selects whether to plot required cases, controls, or total sample size. `compare_tests = TRUE` plots both tests together.

Supported `x_var` values are the same heterogeneity and misclassification variables documented for `plot_cc_power()`, plus "power" for target power. Fixed sample-size variables such as "N_case" are not valid because sample size is the MSSN output.

Multiplier behavior matches `plot_cc_power()`: `pheno_error_multiplier` multiplies baseline `theta_base` and `phi_base`; `geno_error_multiplier` multiplies the corresponding baseline genotype-error parameters for the selected genotype misclassification model. All arguments in `...` remain fixed while `x_var` is swept. The y-axis is the required number of selected cases, controls, or total individuals returned by `cc_mssn()`, according to `sample_size`. For locus-heterogeneity sweeps, an exact $\pi = 0$ design has no finite MSSN when target power exceeds α . That scientifically structural boundary is retained with MSSN NA, `finite_mssn = FALSE`, and `status = "no finite MSSN"`; other errors are propagated unchanged.

Value

A ggplot object, or a data frame if `return_data = TRUE`.

See Also

`cc_mssn`, `plot_cc_power`.

Examples

```
plot_cc_mssn(
  x_var = "geno_error_multiplier",
  x_values = c(0, 1, 2),
  test = "trend",
  input_mode = "model_based",
  power = 0.80,
  alpha = 0.05,
  prev = 0.10,
  pd = 0.25,
  R2 = 2,
  MOI = "M",
  geno_misclass = "3p",
  e01_base = 0.02,
  e02_base = 0.01,
  e03_base = 0.005,
  k = 1
)
```

plot_cc_power

Plot Case-Control Power

Description

Sweeps one x-axis parameter and repeatedly calls `cc_power()` to plot case-control power. The wrapper supports both `input_mode = "model_based"` and `input_mode = "model_free"`.

Usage

```
plot_cc_power(
  x_var,
  x_values,
  test = c("genotypes", "trend"),
  input_mode = c("model_based", "model_free"),
  compare_tests = FALSE,
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

<code>x_var</code>	Character. Parameter to vary on the x-axis.
<code>x_values</code>	Numeric vector of x-axis values.
<code>test</code>	One of "genotypes" or "trend".
<code>input_mode</code>	One of "model_based" or "model_free".
<code>compare_tests</code>	Logical. If TRUE, plot genotype and trend tests together and ignore test.
<code>title</code>	Optional character title override.
<code>x_label</code>	Optional x-axis label override.
<code>y_label</code>	Optional y-axis label override.
<code>return_data</code>	Logical. If TRUE, return the data frame instead of a ggplot.
<code>...</code>	Arguments passed to <code>cc_power()</code> .

Details

Supported `x_var` values include fixed sample-size variables "N_case", "N_ctrl", and "N_total"; model and design variables "alpha", "prev", "pd", "R2", and "k"; locus-heterogeneity variables "pi" and "locus_het_rate"; phenotype-misclassification variables "theta", "phi", and "pheno_error_multiplier"; and genotype-misclassification variables "e", "e1", "e2", "e01", "e02", "e03", case/control differential three-parameter error rates, "geno_error_multiplier", and "diff_multiplier".

When `x_var = "pheno_error_multiplier"`, baseline phenotype-error values are read from `theta_base` and `phi_base`, falling back to `theta` and `phi` if the baseline arguments are not supplied. When `x_var = "geno_error_multiplier"`, baseline genotype-error values are read from `e_base`; `e1_base` and `e2_base`; `e01_base`, `e02_base`, and `e03_base`; or case/control baseline parameters for `geno_misclass = "diff3p"`. All arguments supplied through `...` remain fixed while `x_var` is swept. The x-axis contains `x_values`; the y-axis is the power returned by `cc_power()` for the selected genotype chi-square or trend test. With `compare_tests = TRUE`, both powers are drawn on the same axes. When sweeping `x_var = "pi"`, set `locus_het = TRUE`. The exact $\pi = 0$ null boundary is retained in returned and plotted data with zero NCP and power equal to α .

Value

A ggplot object, or a data frame if return_data = TRUE.

See Also

[cc_power](#), [plot_cc_mssn](#).

Examples

```
g_aff <- c((1 - 0.05)^2, 2 * 0.05 * (1 - 0.05), 0.05^2)
g_unaff <- c((1 - 0.15)^2, 2 * 0.15 * (1 - 0.15), 0.15^2)

plot_cc_power(
  x_var = "phi",
  x_values = c(0, 0.02, 0.04),
  test = "genotypes",
  input_mode = "model_free",
  N_case = 250,
  alpha = 0.01,
  g1 = g_aff,
  g0 = g_unaff,
  prev = 0.05,
  pheno_misclass = TRUE,
  theta = 0,
  k = 1
)
```

plot_ngs_mssn

Plot MSSN Sensitivity for Sequencing Designs

Description

Plots analytic minimum sample size necessary (MSSN) against equal fixed sequencing coverage for either a case-control sequencing trend design or a TDT1-NGS design. Calculations delegate to the corresponding public MSSN API.

Usage

```
plot_ngs_mssn(
  design = c("cc", "tdt"),
  coverage,
  seq_error,
  return_data = FALSE,
  ...
)
```

Arguments

design	Either "cc" for <code>cc_ngs_mssn()</code> or "tdt" for <code>tdt_ngs_mssn()</code> .
coverage	Non-empty numeric vector of integer fixed sequencing depths. TDT1-NGS requires coverage of at least 2.
seq_error	Non-empty numeric vector of symmetric per-read sequencing error probabilities in $[0, 0.5)$. Multiple values produce separate curves.
return_data	Logical. If TRUE, return the exact plotting data rather than a ggplot object.
...	Fixed arguments passed to <code>cc_ngs_mssn()</code> or <code>tdt_ngs_mssn()</code> . For case-control designs, <code>locus_het = TRUE</code> permits a vector of π values. TDT1-NGS accepts a fixed <code>locus_het</code> and <code>heter_rate</code> through ...; its public convention is <code>heter_rate = 1 - pi</code> .

Details

The primary case-control y-axis is required cases; returned plot data also retain required controls, total MSSN, and achieved power. The primary TDT1-NGS y-axis is required complete trios; returned data also retain total individuals and achieved power.

At the case-control exact-null boundary $\pi = 0$, no finite MSSN exists when requested power exceeds alpha. The row is retained with MSSN and achieved power set to NA, `finite_mssn = FALSE`, and status "no finite MSSN". No artificial finite value is plotted. Only this expected scientific condition is converted; invalid inputs and unexpected calculation errors propagate unchanged.

Coverage is fixed and equal for the relevant study members. These plots do not model variable/BGE coverage distributions, cost optimization, or sample-specific depth. They use no simulation.

Value

A ggplot object with exact results in `plot$data`, or a data frame when `return_data = TRUE`.

See Also

`plot_ngs_power`, `cc_ngs_mssn`, `tdt_ngs_mssn`

Examples

```
plot_ngs_mssn(
  design = "tdt", coverage = c(4, 12, 20), seq_error = 0.005,
  power = 0.80, pd = 0.325, R1 = 1.2, alpha = 5e-8,
  verbose = FALSE
)
```

plot_ngs_power

*Plot Power Sensitivity for Sequencing Designs***Description**

Plots prospective analytic power against equal fixed sequencing coverage for either a case-control sequencing trend design or a TDT1-NGS design. Exact values are obtained by repeated calls to the corresponding validated public design function.

Usage

```
plot_ngs_power(
  design = c("cc", "tdt"),
  coverage,
  seq_error,
  target_power = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

design	Either "cc" for <code>cc_ngs_power()</code> or "tdt" for <code>tdt_ngs_power()</code> .
coverage	Non-empty numeric vector of integer fixed sequencing depths. TDT1-NGS requires coverage of at least 2.
seq_error	Non-empty numeric vector of symmetric per-read sequencing error probabilities in $[0, 0.5)$. Multiple values produce separate curves.
target_power	Optional finite scalar in $(0, 1)$. If supplied, a dashed horizontal reference line is added without changing calculations.
return_data	Logical. If TRUE, return the exact plotting data rather than a ggplot object.
...	Fixed arguments passed to <code>cc_ngs_power()</code> or <code>tdt_ngs_power()</code> . For case-control designs, <code>locus_het = TRUE</code> permits a vector of π values. TDT1-NGS accepts a fixed <code>locus_het</code> and <code>heter_rate</code> through ...; its public convention is <code>heter_rate = 1 - pi</code> .

Details

Case-control output retains coverage, sequencing error, locus-homogeneity fraction, sample sizes, NCP, and power. At $\pi = 0$, the exact null point is retained with NCP zero and power equal to α . TDT1-NGS uses raw read-count probabilities and does not use the case-control hard-call model.

Coverage is fixed and equal for the relevant study members. These plots do not model variable/BGE coverage distributions, cost optimization, or sample-specific depth. They introduce no simulation and delegate all statistical calculations to public PAWEH sequencing design APIs.

Value

A ggplot object with exact results in plot\$data, or a data frame when return_data = TRUE.

See Also

[plot_ngs_mssn](#), [cc_ngs_power](#), [tdt_ngs_power](#)

Examples

```
plot_ngs_power(
  design = "cc", coverage = c(4, 12, 20), seq_error = c(0, 0.01),
  N_case = 1000, alpha = 0.05, prev = 0.05, pd = 0.30, R2 = 1.8,
  MOI = "M", verbose = FALSE
)
```

plot_qtl_anova_mssn	<i>Plot One-Way ANOVA Minimum Sample Size Under the Falconer Model</i>
---------------------	--

Description

Sweeps one model or design parameter and repeatedly calls [qtl_anova_mssn\(\)](#).

Usage

```
plot_qtl_anova_mssn(
  x_var,
  x_values,
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

x_var	Parameter to vary. One of "power", "alpha", "qtl_var", "tau", or "pd".
x_values	Numeric vector of at least two finite x-axis values.
title	Optional plot-title override.
x_label	Optional x-axis-label override.
y_label	Optional y-axis-label override.
return_data	Logical. If TRUE, returns plotting data rather than a ggplot object.
...	Fixed arguments passed to qtl_anova_mssn() .

Details

The x-axis is `x_values` for the selected `x_var`; the y-axis is the minimum total sample size returned by `qtl_anova_mssn()`. Target power and other arguments in `...` remain fixed unless selected as `x_var`.

Value

A ggplot object, or a data frame with the swept parameter and `required_N` if `return_data = TRUE`.

See Also

[qtl_anova_mssn](#), [plot_qtl_anova_power](#).

Examples

```
plot_qtl_anova_mssn(  
  x_var = "qtl_var", x_values = c(0.015, 0.025, 0.05),  
  power = 0.8, alpha = 0.0001, tau = 0.5, pd = 0.15  
)
```

`plot_qtl_anova_power` *Plot One-Way ANOVA Power Under the Falconer Model*

Description

Sweeps one model or design parameter and repeatedly calls [qtl_anova_power\(\)](#) for a single continuous quantitative trait.

Usage

```
plot_qtl_anova_power(  
  x_var,  
  x_values,  
  title = NULL,  
  x_label = NULL,  
  y_label = NULL,  
  return_data = FALSE,  
  ...  
)
```

Arguments

x_var	Parameter to vary. One of "N", "alpha", "qtl_var", "tau", or "pd".
x_values	Numeric vector of at least two finite x-axis values.
title	Optional plot-title override.
x_label	Optional x-axis-label override.
y_label	Optional y-axis-label override.
return_data	Logical. If TRUE, returns plotting data rather than a ggplot object.
...	Fixed arguments passed to qtl_anova_power().

Details

The x-axis is x_values for the selected x_var; the y-axis is one-way ANOVA power from qtl_anova_power(). Every argument in ... remains fixed while the selected parameter is swept.

Value

A ggplot object, or a data frame with the swept parameter and power if return_data = TRUE.

See Also

[qtl_anova_power](#), [plot_qtl_anova_mssn](#).

Examples

```
plot_qtl_anova_power(
  x_var = "N", x_values = c(600, 800, 1000),
  alpha = 0.0001, qtl_var = 0.025, tau = 0.5, pd = 0.15
)
```

plot_qtl_genotype_distribution

Plot Falconer Genotype-Specific Quantitative-Trait Distributions

Description

Displays the three normal quantitative-trait distributions implied by a validated single-trait Falconer model. Large QTL variance separates the genotype distributions relative to their residual variation; small QTL variance makes them overlap heavily.

Usage

```
plot_qtl_genotype_distribution(
  qtl_var,
  tau,
  pd,
  type = c("density", "histogram", "binned"),
  scale = c("density", "frequency"),
  n = 3000,
  seed = NULL,
  show_means = TRUE,
  verbose = FALSE,
  title = NULL,
  return_data = FALSE,
  bins = 30L
)
```

Arguments

qtl_var	QTL variance in (0, 1).
tau	Dominance-to-additivity ratio.
pd	Increaser-allele frequency in (0, 1).
type	One of "density" for theoretical normal curves, "histogram" for a simulated population, or "binned" for a deterministic histogram-style display calculated from Normal CDF differences.
scale	Either "density" or "frequency". For theoretical curves, frequency scaling weights each density by its Hardy–Weinberg genotype frequency. For histograms, it selects normalized densities or counts.
n	Total population size simulated in histogram mode.
seed	Optional non-negative integer used locally for reproducible histogram simulation. The caller's random-number state is restored.
show_means	Logical; mark the three theoretical genotype means.
verbose	Logical; passed to the validated Falconer parameter backend.
title	Optional plot title.
return_data	Logical; return the plotting data frame instead of the ggplot object. The Falconer model is retained in its falconer_model attribute.
bins	Positive integer number of equal-width bins used when type = "binned". This does not affect the existing density or simulated histogram modes.

Details

Conditional on genotype j , the plotted trait follows a normal distribution with the corresponding Falconer mean and common residual variance $1 - \text{qtl_var}$. Genotypes 0, 1, and 2 have Hardy–Weinberg frequencies $(1 - \text{pd})^2$, $2 * \text{pd} * (1 - \text{pd})$, and pd^2 . Genotypes may also be interpreted as bb, Bb, and BB, respectively.

Density and binned modes are analytic and perform no simulation. In binned mode, each conditional bin probability is the difference between Normal CDF values at the upper and lower bin boundaries. Histogram mode first samples genotypes from their Hardy–Weinberg frequencies and then samples trait values from the corresponding conditional normal distributions. The x-axis is the quantitative-trait value. In analytic density mode, the y-axis is genotype-specific density when `scale = "density"` and population-weighted density when `scale = "frequency"`. Binned mode uses conditional bin probability divided by bin width, with an additional genotype-frequency multiplier under `scale = "frequency"`. In simulated histogram mode, the y-axis is normalized density or bin count for those respective scales.

Value

A ggplot object, or a plotting data frame when `return_data = TRUE`.

See Also

[qtl_falconer_parameters\(\)](#), [plot_qtl_anova_power\(\)](#), and [plot_qtl_threshold_chisq_power\(\)](#).

Examples

```
# A large QTL effect visibly separates the genotype distributions.
plot_qtl_genotype_distribution(
  qtl_var = 0.5, tau = 0, pd = 0.25, scale = "frequency"
)

# A very small QTL effect produces almost complete overlap.
plot_qtl_genotype_distribution(
  qtl_var = 0.0005, tau = 0, pd = 0.25, scale = "frequency"
)
```

plot_qtl_threshold_chisq_mssn

Plot Threshold-Selected Genotype Chi-Square Minimum Sample Size

Description

Sweeps one model, selection, or design parameter and repeatedly calls [qtl_threshold_chisq_mssn\(\)](#).

Usage

```
plot_qtl_threshold_chisq_mssn(
  x_var,
  x_values,
  sample_size = c("total", "case", "control"),
  title = NULL,
  x_label = NULL,
  y_label = NULL,
```

```

    return_data = FALSE,
    ...
)

```

Arguments

<code>x_var</code>	Parameter to vary. One of "power", "alpha", "qtl_var", "tau", "pd", "x_upper", "x_lower", or "k".
<code>x_values</code>	Numeric vector of at least two finite x-axis values.
<code>sample_size</code>	Selected-sample result to plot: "total", "case", or "control". Population screening counts are deliberately excluded because they are not statistical MSSN values.
<code>title</code>	Optional plot-title override.
<code>x_label</code>	Optional x-axis-label override.
<code>y_label</code>	Optional y-axis-label override.
<code>return_data</code>	Logical. If TRUE, returns plotting data rather than a ggplot object.
<code>...</code>	Fixed arguments passed to <code>qtl_threshold_chisq_mssn()</code> .

Details

The x-axis is `x_values` for the selected model, selection, or design parameter. The y-axis is the selected case, control, or total MSSN requested by `sample_size`; it is not a source-population screening count. Other arguments remain fixed while `x_var` is swept.

Value

A ggplot object, or a data frame with the swept parameter and the selected MSSN result if `return_data = TRUE`.

See Also

[qtl_threshold_chisq_mssn](#), [plot_qtl_threshold_chisq_power](#).

Examples

```

plot_qtl_threshold_chisq_mssn(
  x_var = "qtl_var", x_values = c(0.015, 0.025, 0.05),
  power = 0.8, alpha = 0.0001, tau = 0.5, pd = 0.15,
  x_upper = 5, x_lower = 5, k = 1
)

```

plot_qtl_threshold_chisq_power

Plot Threshold-Selected Genotype Chi-Square Power

Description

Sweeps one model, selection, or design parameter and repeatedly calls `qtl_threshold_chisq_power()`.

Usage

```
plot_qtl_threshold_chisq_power(
  x_var,
  x_values,
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

<code>x_var</code>	Parameter to vary. One of "N_case", "alpha", "qtl_var", "tau", "pd", "x_upper", "x_lower", or "k".
<code>x_values</code>	Numeric vector of at least two finite x-axis values.
<code>title</code>	Optional plot-title override.
<code>x_label</code>	Optional x-axis-label override.
<code>y_label</code>	Optional y-axis-label override.
<code>return_data</code>	Logical. If TRUE, returns plotting data rather than a ggplot object.
<code>...</code>	Fixed arguments passed to <code>qtl_threshold_chisq_power()</code> .

Details

The x-axis is `x_values` for the selected model, selection, or design parameter. The y-axis is power for the two-df genotype chi-square test after upper-tail case and lower-tail control selection. Arguments in `...` remain fixed while `x_var` is swept.

Value

A ggplot object, or a data frame with the swept parameter and power if `return_data = TRUE`.

See Also

[qtl_threshold_chisq_power](#), [plot_qtl_threshold_chisq_mssn](#).

Examples

```
plot_qtl_threshold_chisq_power(
  x_var = "N_case", x_values = c(75, 100, 125, 150),
  alpha = 0.0001, qtl_var = 0.025, tau = 0.5, pd = 0.15,
  x_upper = 5, x_lower = 5, k = 1
)
```

plot_tdt_mssn

Plot TDT Minimum Sample Size Necessary

Description

Sweeps one x-axis parameter and repeatedly calls `tdt_mssn()` to plot MSSN in affected trios. The wrapper supports both `input_mode = "model_based"` and `input_mode = "model_free"`.

Usage

```
plot_tdt_mssn(
  x_var,
  x_values,
  scenario = c("auto", "no_error", "misclassification", "heterogeneity"),
  input_mode = c("model_based", "model_free"),
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

<code>x_var</code>	Character. Parameter to vary on the x-axis.
<code>x_values</code>	Numeric vector of x-axis values.
<code>scenario</code>	One of "auto", "no_error", "misclassification", or "heterogeneity".
<code>input_mode</code>	One of "model_based" or "model_free".
<code>title</code>	Optional character title override.
<code>x_label</code>	Optional x-axis label override.
<code>y_label</code>	Optional y-axis label override.
<code>return_data</code>	Logical. If TRUE, return the data frame instead of a ggplot.
<code>...</code>	Arguments passed to <code>tdt_mssn()</code> .

Details

scenario selects the component of the canonical TDT backend output: "no_error", "misclassification", or "heterogeneity". With scenario = "auto", the wrapper chooses the scenario from x_var using the same rules as [plot_tdt_power\(\)](#). This wrapper does not include a compare-scenarios mode. x_var = "N" is not allowed because required trios are the output; use "target_power" to sweep target power.

With input_mode = "model_based" (the default), this wrapper's x_var choices and output are numerically identical to sweeping [tdt_mssn\(\)](#) directly. "pd", "prev", "R1", "R2", and "delta_prime" are only valid x_var choices in this mode.

With input_mode = "model_free", sweep "ET", "ENT", or "n_trios" instead (or any of the shared "misclass_rate", "heter_rate", "locus_het_rate", or "pheno_error_multiplier" variables). A resolved scenario = "heterogeneity" requires a fixed pd argument, and scenario = "misclassification" requires fixed pd and prev arguments, because the conditional backend needs them to apply heter_rate/misclass_rate to the supplied ET/ENT. Unlike [tdt_mssn\(\)](#)'s own default of 0.01, heter_rate and misclass_rate default to 0 here in input_mode = "model_free" (unless fixed via ... or swept via x_var), so scenario = "no_error" always works without pd/prev. All arguments passed through ... remain fixed while x_var is swept. The y-axis is MSSN expressed as the required number of affected trios for the resolved scenario. When x_var = "heter_rate", it is the heterogeneous trio fraction $1 - \pi$; target_power remains fixed unless it is itself the swept variable.

Value

A ggplot object, or a data frame if return_data = TRUE.

See Also

[tdt_mssn](#), [plot_tdt_power](#), [plot_tdt_mssn_phenotype_misclassification](#), and [plot_tdt_mssn_locus_heterogene](#)

Examples

```
plot_tdt_mssn(
  x_var = "misclass_rate",
  x_values = c(0, 0.02, 0.04),
  scenario = "misclassification",
  input_mode = "model_based",
  target_power = 0.80,
  pd = 0.30,
  prev = 0.05,
  R1 = 1.5,
  R2 = 2.25,
  alpha = 0.05,
  delta_prime = 1
)
```

plot_tdt_mssn_locus_heterogeneity

Plot TDT MSSN Sensitivity to Locus Heterogeneity

Description

Plot required number of trios vs heterogeneity rate ($1 - \pi$), holding misclassification fixed.

Usage

```
plot_tdt_mssn_locus_heterogeneity(
  pd,
  prev,
  R1,
  R2,
  alpha = 0.05,
  delta_prime = 1,
  target_power,
  heter_seq = seq(0, 0.5, by = 0.05),
  misclass_fixed = 0,
  title = "Required trios vs heterogeneity ( $1 - \pi$ )"
)
```

Arguments

pd	Disease allele frequency.
prev	Disease prevalence.
R1	Genotype relative risk for heterozygotes.
R2	Genotype relative risk for risk homozygotes.
alpha	Significance level.
delta_prime	Linkage disequilibrium scaling parameter.
target_power	Desired statistical power.
heter_seq	Numeric vector. Sequence of heterogeneity rates.
misclass_fixed	Numeric. Misclassification rate held fixed (default 0). (Other params same meaning as in plot_tdt_power_phenotype_misclassification.)
title	Plot title.

Details

The x-axis is the heterogeneous trio fraction $1 - \pi$; the homogeneous/linked fraction is π . The y-axis is MSSN, the required number of affected trios returned by tdt_mssn() for fixed target_power. misclass_fixed and genetic-model parameters remain fixed.

Value

A ggplot object, returned invisibly after it is printed.

See Also

[tdt_mssn](#), [plot_tdt_mssn_phenotype_misclassification](#), and [plot_tdt_mssn](#).

Examples

```
# Required sample size sensitivity to locus heterogeneity
plot_tdt_mssn_locus_heterogeneity(
  target_power = 0.80,
  pd = 0.30,
  prev = 0.05,
  R1 = 1.5,
  R2 = 2.25,
  heter_seq = c(0, 0.1, 0.2),
  misclass_fixed = 0
)
```

plot_tdt_mssn_phenotype_misclassification

Plot TDT MSSN Sensitivity to Phenotype Misclassification

Description

Plot required number of trios vs misclassification rate (π_01), holding heterogeneity fixed.

Usage

```
plot_tdt_mssn_phenotype_misclassification(
  pd,
  prev,
  R1,
  R2,
  alpha = 0.05,
  delta_prime = 1,
  target_power,
  misclass_seq = seq(0, 0.15, by = 0.01),
  heter_fixed = 0,
  title = "Required trios vs misclassification ( $\pi_01$ )"
)
```

Arguments

<code>pd</code>	Disease allele frequency.
<code>prev</code>	Disease prevalence.
<code>R1</code>	Genotype relative risk for heterozygotes.
<code>R2</code>	Genotype relative risk for risk homozygotes.
<code>alpha</code>	Significance level.
<code>delta_prime</code>	Linkage disequilibrium scaling parameter.
<code>target_power</code>	Numeric. Desired power for sample size calculation.
<code>misclass_seq</code>	Numeric vector. Sequence of misclassification rates.
<code>heter_fixed</code>	Numeric. Heterogeneity rate held fixed (default 0). (Other params same meaning as in <code>plot_tdt_power_phenotype_misclassification</code> .)
<code>title</code>	Plot title.

Details

The x-axis is phenotype misclassification probability π_{01} . The y-axis is MSSN, the required number of affected trios returned by `tdt_mssn()` for fixed `target_power`. `heter_fixed` and all genetic-model parameters remain fixed.

Value

A ggplot object, returned invisibly after it is printed.

See Also

[tdt_mssn](#), [plot_tdt_mssn_locus_heterogeneity](#), and [plot_tdt_mssn](#).

Examples

```
# Required sample size sensitivity to phenotype misclassification
plot_tdt_mssn_phenotype_misclassification(
  target_power = 0.80,
  pd = 0.30,
  prev = 0.05,
  R1 = 1.5,
  R2 = 2.25,
  misclass_seq = c(0, 0.03, 0.06),
  heter_fixed = 0
)
```

plot_tdt_power	<i>Plot TDT Power</i>
----------------	-----------------------

Description

Sweeps one x-axis parameter and repeatedly calls `tdt_power()` to plot Transmission Disequilibrium Test (TDT) power. The wrapper supports both `input_mode = "model_based"` and `input_mode = "model_free"`.

Usage

```
plot_tdt_power(
  x_var,
  x_values,
  scenario = c("auto", "no_error", "misclassification", "heterogeneity"),
  input_mode = c("model_based", "model_free"),
  title = NULL,
  x_label = NULL,
  y_label = NULL,
  return_data = FALSE,
  ...
)
```

Arguments

<code>x_var</code>	Character. Parameter to vary on the x-axis.
<code>x_values</code>	Numeric vector of x-axis values.
<code>scenario</code>	One of "auto", "no_error", "misclassification", or "heterogeneity".
<code>input_mode</code>	One of "model_based" or "model_free".
<code>title</code>	Optional character title override.
<code>x_label</code>	Optional x-axis label override.
<code>y_label</code>	Optional y-axis label override.
<code>return_data</code>	Logical. If TRUE, return the data frame instead of a ggplot.
<code>...</code>	Arguments passed to <code>tdt_power()</code> .

Details

`scenario` selects the component of the canonical TDT backend output: "no_error", "misclassification", or "heterogeneity". With `scenario = "auto"`, the wrapper chooses misclassification for `x_var = "misclass_rate"` or `"pheno_error_multiplier"`, heterogeneity for `x_var = "heter_rate"` or `"locus_het_rate"`, and no error otherwise. This wrapper does not include a compare-scenarios mode.

With `input_mode = "model_based"` (the default), this wrapper's `x_var` choices and output are numerically identical to sweeping `tdt_power()` directly. "pd", "prev", "R1", "R2", and "delta_prime" are only valid `x_var` choices in this mode.

With `input_mode = "model_free"`, sweep "ET" or "ENT" instead (or any of the shared "misclass_rate", "heter_rate", "locus_het_rate", or "pheno_error_multiplier" variables). A resolved scenario = "heterogeneity" requires a fixed `pd` argument, and `scenario = "misclassification"` requires fixed `pd` and `prev` arguments, because the conditional backend needs them to apply `heter_rate/misclass_rate` to the supplied ET/ENT. Unlike `tdt_power()`'s own default of 0.01, `heter_rate` and `misclass_rate` default to 0 here in `input_mode = "model_free"` (unless fixed via ... or swept via `x_var`), so `scenario = "no_error"` always works without `pd/prev`. All arguments passed through ... remain fixed while `x_var` is swept. The x-axis contains `x_values`; the y-axis is TDT power for the resolved scenario. When `x_var = "heter_rate"`, the axis is the heterogeneous trio fraction $1 - \pi$.

Value

A ggplot object, or a data frame if `return_data = TRUE`.

See Also

[tdt_power](#), [plot_tdt_mssn](#), [plot_tdt_power_phenotype_misclassification](#), and [plot_tdt_power_locus_heterogeneity](#)

Examples

```
plot_tdt_power(
  x_var = "heter_rate",
  x_values = c(0, 0.1, 0.2),
  scenario = "heterogeneity",
  input_mode = "model_based",
  N = 600,
  pd = 0.30,
  prev = 0.05,
  R1 = 1.5,
  R2 = 2.25,
  alpha = 0.05,
  delta_prime = 1
)
```

plot_tdt_power_locus_heterogeneity

Plot TDT Power Sensitivity to Locus Heterogeneity

Description

Plot TDT power vs locus heterogeneity rate ($1 - \pi$), holding misclassification fixed.

Usage

```
plot_tdt_power_locus_heterogeneity(
  pd,
  prev,
  R1,
```

```

    R2,
    alpha = 0.05,
    delta_prime = 1,
    N,
    heter_seq = seq(0, 0.5, by = 0.05),
    misclass_fixed = 0,
    title = "TDT power vs heterogeneity (1 - pi)"
  )

```

Arguments

pd	Disease allele frequency.
prev	Disease prevalence.
R1	Genotype relative risk for heterozygotes.
R2	Genotype relative risk for risk homozygotes.
alpha	Significance level.
delta_prime	Linkage disequilibrium scaling parameter.
N	Number of affected trios.
heter_seq	Numeric vector. Sequence of heterogeneity rates.
misclass_fixed	Numeric. Misclassification rate held fixed (default 0). (Other params same meaning as in plot_tdt_power_phenotype_misclassification.)
title	Plot title.

Details

The x-axis is the heterogeneous trio fraction $1 - \pi$; the homogeneous/linked fraction is π . The y-axis is heterogeneity-scenario power from `tdt_power()`. `N`, `misclass_fixed`, and all genetic-model parameters remain fixed while `heter_seq` is swept.

Value

A ggplot object, returned invisibly after it is printed.

See Also

[tdt_power](#), [plot_tdt_power_phenotype_misclassification](#), and [plot_tdt_power](#).

Examples

```

# Power sensitivity to locus heterogeneity
plot_tdt_power_locus_heterogeneity(
  N      = 600,
  pd     = 0.30,
  prev   = 0.05,
  R1     = 1.5,
  R2     = 2.25,
  heter_seq = c(0, 0.1, 0.2),
  misclass_fixed = 0
)

```

)

plot_tdt_power_phenotype_misclassification

Plot TDT Power Sensitivity to Phenotype Misclassification

Description

Plot TDT power vs phenotype misclassification rate (π_{01}), holding heterogeneity fixed.

Usage

```
plot_tdt_power_phenotype_misclassification(
  pd,
  prev,
  R1,
  R2,
  alpha = 0.05,
  delta_prime = 1,
  N,
  misclass_seq = seq(0, 0.15, by = 0.01),
  heter_fixed = 0,
  title = "TDT power vs misclassification ( $\pi_{01}$ )"
)
```

Arguments

pd	Numeric in (0,1). High-risk allele frequency at the marker.
prev	Numeric in (0,1). Disease prevalence (ϕ_1).
R1, R2	Numeric. Genotype relative risks.
alpha	Numeric. Significance level (default 0.05).
delta_prime	Numeric. LD scale parameter (default 1).
N	Integer. Number of affected trios.
misclass_seq	Numeric vector. Sequence of misclassification rates.
heter_fixed	Numeric. Heterogeneity rate held fixed (default 0).
title	Character. Plot title.

Details

The x-axis is phenotype misclassification probability π_{01} and the y-axis is the corresponding power from `tdt_power()`. N is a fixed number of affected trios; `heter_fixed` and all genetic-model parameters remain fixed while `misclass_seq` is swept.

Value

A ggplot object, returned invisibly after it is printed.

See Also

[tdt_power](#), [plot_tdt_power_locus_heterogeneity](#), and [plot_tdt_power](#).

Examples

```
# Power sensitivity to phenotype misclassification
plot_tdt_power_phenotype_misclassification(
  N      = 600,
  pd     = 0.30,
  prev   = 0.05,
  R1     = 1.5,
  R2     = 2.25,
  misclass_seq = c(0, 0.03, 0.06),
  heter_fixed  = 0
)
```

plot_tdt_surface3d	<i>Plot a Generalized 3D TDT Sensitivity Surface</i>
--------------------	--

Description

Builds an interactive sensitivity surface for model-based TDT power or minimum sample size. Any two supported model parameters can be swept while the remaining parameters are held fixed. Every grid point is evaluated by the canonical [tdt_power\(\)](#) or [tdt_mssn\(\)](#) backend.

Usage

```
plot_tdt_surface3d(
  metric = c("power", "mssn"),
  scenario = c("misclassification", "heterogeneity", "no_error"),
  x = "pd",
  y = "misclass_rate",
  x_values = NULL,
  y_values = NULL,
  N = 600,
  target_power = 0.8,
  pd = 0.3,
  prev = 0.05,
  R1 = 1.5,
  R2 = 2.25,
  alpha = 0.05,
  delta_prime = 1,
  misclass_rate = 0.01,
```

```

heter_rate = 0.1,
ceiling_N = TRUE,
title = NULL
)

```

Arguments

metric	Character. Surface height: "power" (default) or "mssn".
scenario	Character. One of "misclassification" (default), "heterogeneity", or "no_error".
x, y	Character scalars naming two distinct swept parameters. Supported axes are "pd", "prev", "R1", "R2", "alpha", "delta_prime", "misclass_rate", and "heter_rate".
x_values, y_values	Numeric grid values for x and y. Each must contain at least two distinct, finite values. If NULL, documented parameter-specific defaults are used; see Details.
N	Numeric scalar greater than zero. Number of affected trios, used when metric = "power".
target_power	Numeric scalar in (0, 1), used when metric = "mssn".
pd	Numeric in (0, 1). Fixed risk-allele frequency when pd is not an axis.
prev	Numeric in (0, 1). Fixed disease prevalence when prev is not an axis.
R1, R2	Positive numeric scalars. Fixed genotype relative risks when they are not axes.
alpha	Numeric in (0, 1). Fixed significance level when alpha is not an axis.
delta_prime	Numeric in [0, 1]. Fixed positive linkage-disequilibrium scale when delta_prime is not an axis.
misclass_rate	Numeric in [0, 1). Fixed phenotype misclassification rate when it is not an axis.
heter_rate	Numeric in [0, 1). Fixed locus-heterogeneity rate when it is not an axis.
ceiling_N	Logical scalar. For MSSN surfaces, whether to plot the integer ceiling of the canonical sample-size result (default TRUE). The unrounded result remains available in the attached surface data.
title	Optional character scalar. If NULL, an informative title is generated from metric and scenario.

Details

The supported axes and their default grids are: `pd = seq(0.10, 0.50, length.out = 20)`, `prev = seq(0.01, 0.10, length.out = 20)`, `R1 = seq(1.1, 2, length.out = 20)`, `R2 = seq(1.2, 3, length.out = 20)`, `alpha = seq(0.01, 0.10, length.out = 20)`, `delta_prime = seq(0.25, 1, length.out = 20)`, `misclass_rate = seq(0, 0.20, length.out = 20)`, and `heter_rate = seq(0, 0.50, length.out = 20)`. The package's implemented LD convention is $D = \text{delta_prime} * \text{pd} * (1 - \text{pd})$, where `delta_prime` is the proportion of maximum positive disequilibrium: 0 represents linkage equilibrium and 1 represents maximum positive LD under the model assumptions. Negative LD would require a different, allele-frequency-dependent normalization and is not represented by this parameterization.

The surface is deliberately model-based. `N` and `target_power` define the design calculation and therefore remain fixed rather than becoming axes. The model-free inputs `ET`, `ENT`, and `n_trios`, as well as nonnumeric controls such as `verbose`, are outside this visualization interface. `MSSN` is measured in affected-child trios. `heter_rate` is the heterogeneous fraction, $1 - \pi$.

The canonical TDT backends report separate scenarios, not a combined-error scenario. Consequently, `misclass_rate` is an active axis only for `scenario = "misclassification"`, and `heter_rate` is active only for `scenario = "heterogeneity"`. An inactive modifier axis is rejected rather than silently producing a flat, misleading surface.

This function requires the optional `plotly` package. The plotting layer introduces no independent statistical approximation: it only constructs the grid, delegates calculations to `tdt_power()` or `tdt_mssn()`, extracts the selected scenario, and renders the result.

Value

An object inheriting from `"plotly"` and `"htmlwidget"`. The Cartesian grid is attached as `attr(x, "surface_data")`, with dynamically named axis columns, `metric_value` (the plotted height), and `raw_metric_value`. The plotted z matrix is attached as `attr(x, "surface_matrix")`, and calculation settings as `attr(x, "surface_spec")`.

See Also

`tdt_power()`, `tdt_mssn()`, `plot_tdt_power()`, `plot_tdt_mssn()`

Examples

```
if (requireNamespace("plotly", quietly = TRUE)) {
  power_surface <- plot_tdt_surface3d(
    metric = "power", scenario = "misclassification",
    x = "pd", y = "misclass_rate",
    x_values = c(0.2, 0.3, 0.4),
    y_values = c(0, 0.03, 0.06),
    N = 600
  )

  mssn_surface <- plot_tdt_surface3d(
    metric = "mssn", scenario = "heterogeneity",
    x = "prev", y = "heter_rate",
    x_values = c(0.02, 0.05, 0.08),
    y_values = c(0, 0.1, 0.2),
    target_power = 0.8
  )
}
```

qtl_anova_mssn	<i>Minimum Sample Size for One-Way ANOVA Under the Falconer Model</i>
----------------	---

Description

Finds the smallest sufficient integer total sample size for the three-group one-way ANOVA. Both denominator degrees of freedom and the non-centrality parameter are recalculated for every candidate sample size.

Usage

```
qtl_anova_mssn(
  power,
  alpha,
  qtl_var,
  tau,
  pd,
  count_method = c("rounded", "expected"),
  multiple_of_three = TRUE,
  verbose = TRUE
)
```

Arguments

power	Target power in (0, 1).
alpha	Significance level in (0, 1).
qtl_var	Variance in the standardized trait attributable to the QTL. Must be in (0, 1).
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in (0, 1).
count_method	Genotype count method. "rounded" rounds expected genotype counts to the nearest integer before computing the ANOVA quantities. "expected" uses expected genotype counts directly without rounding.
multiple_of_three	Logical. If TRUE, searches sample sizes in multiples of three. If FALSE, searches every integer sample size.
verbose	Logical. If TRUE, prints a concise summary.

Details

For every candidate integer N, the genotype counts, denominator degrees of freedom, noncentrality parameter, critical F value, and achieved power are recomputed. With `count_method = "rounded"`, candidates for which any genotype group has zero members are excluded. The first candidate attaining power is returned; setting `multiple_of_three = TRUE` restricts the search grid rather than changing the ANOVA formula. The model and NCP follow textbook Eqs. 6.7–6.8 (p. 328).

Value

An object of class "qtl_anova_mssn" containing the smallest sufficient N, achieved power, degrees of freedom, non-centrality parameter, genotype counts, and Falconer parameters.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.1.1, Eqs. 6.7–6.8, p. 328. doi:[10.1007/9783030611217](https://doi.org/10.1007/9783030611217).

See Also

[qtl_anova_power](#) and [qtl_falconer_parameters](#).

Examples

```
qtl_anova_mssn(
  power = 0.8, alpha = 0.0001, qtl_var = 0.025,
  tau = 0.5, pd = 0.15, verbose = FALSE
)
```

qtl_anova_power

Power for One-Way ANOVA Under the Falconer Model

Description

Calculates power for a one-way ANOVA comparing the means of the three SNP genotype groups for one continuous quantitative trait.

Usage

```
qtl_anova_power(
  N,
  alpha,
  qtl_var,
  tau,
  pd,
  count_method = c("rounded", "expected"),
  verbose = TRUE
)
```

Arguments

N	Integer total sample size greater than 3.
alpha	Significance level in (0, 1).
qtl_var	Variance in the standardized trait attributable to the QTL. Must be in (0, 1).
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in (0, 1).
count_method	Genotype count method. "rounded" rounds expected genotype counts to the nearest integer before computing the ANOVA quantities. "expected" uses expected genotype counts directly without rounding.
verbose	Logical. If TRUE, prints a concise summary.

Details

The three genotype groups have the means returned by `qtl_falconer_parameters()` and a common residual variance $1 - V_{QTL}$. Genotype group counts are either rounded Hardy–Weinberg expectations or unrounded expected counts according to `count_method`. The between-group effect gives the noncentrality parameter in textbook Eq. 6.8, and power is the upper tail of a noncentral F distribution with numerator degrees of freedom 2 and denominator degrees of freedom $N - 3$. Genotype counts follow Eq. 6.7 (Chapter 6, Section 6.1.1, p. 328).

Value

An object of class "qtl_anova_power" containing power, sample size, degrees of freedom, critical value, non-centrality parameter, genotype counts, genotype means, residual variance, and the complete Falconer parameter object.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.1.1, Eqs. 6.7–6.8, p. 328. [doi:10.1007/9783030611217](https://doi.org/10.1007/9783030611217).

See Also

[qtl_anova_mssn](#) and [qtl_falconer_parameters](#).

Examples

```
qtl_anova_power(
  N = 996, alpha = 0.0001, qtl_var = 0.025,
  tau = 0.5, pd = 0.15, verbose = FALSE
)
```

qtl_falconer_parameters

Falconer Parameters for a Single Quantitative Trait

Description

Calculates genotype-specific normal-mixture parameters for a standardized continuous trait under the single-locus Falconer model.

Usage

```
qtl_falconer_parameters(qtl_var, tau, pd, verbose = TRUE)
```

Arguments

qtl_var	Variance in the standardized trait attributable to the QTL. Must be in $(0, 1)$.
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in $(0, 1)$.
verbose	Logical. If TRUE, prints a formatted parameter summary. Set to FALSE to suppress all console output.

Details

Genotypes are ordered as zero, one, and two copies of the increaser allele. Their Hardy-Weinberg mixing proportions are $((1 - p_d)^2, 2p_d(1 - p_d), p_d^2)$. The population-centering constant makes the genotype-weighted trait mean zero, and the residual variance is $1 - V_{QTL}$.

This is a classical, Falconer-style quantitative-genetic parameterization as formulated for this study-design framework by Gordon, Finch, and Kim (2020), Chapter 6, Section 6.1. The population mixture is Eq. 6.1 (p. 324); the additive effect, dominance effect, and centering formulas on p. 325 are unnumbered. The function name does not assert that every exact implemented expression is a numbered equation in Falconer and Mackay.

Value

An object of class "qtl_falconer_parameters". The list contains qtl_var, tau, pd, p_plus, additive effect a, dominance effect delta, centering constant m, genotype means mu, mixing proportions pi, residual_variance, residual_sd, weighted_mean, and total_variance.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.1, pp. 324–325. doi:10.1007/9783030611217.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399–433. doi:10.1017/S0080456800012163.
- Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th ed. Longman.

See Also

[qtl_falconer_threshold_parameters](#), [qtl_anova_power](#), and [qtl_anova_mssn](#).

Examples

```
qtl_falconer_parameters(qtl_var = 0.025, tau = 0.5, pd = 0.15)
```

qtl_falconer_threshold_parameters

Threshold-Selected Parameters Under the Falconer Model

Description

Converts the genotype-specific Falconer trait distributions into upper-tail case and lower-tail control penetrances and conditional genotype frequencies.

Usage

```
qtl_falconer_threshold_parameters(  
  qtl_var,  
  tau,  
  pd,  
  threshold_mode = c("percentile", "direct"),  
  x_upper = NULL,  
  x_lower = NULL,  
  upper_threshold = NULL,  
  lower_threshold = NULL,  
  verbose = TRUE  
)
```

Arguments

qtl_var	Variance in the standardized trait attributable to the QTL. Must be in (0, 1).
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in (0, 1).
threshold_mode	Either "percentile" or "direct".
x_upper	Percentage selected from the upper population tail in percentile mode.
x_lower	Percentage selected from the lower population tail in percentile mode.
upper_threshold	Direct upper-tail threshold in direct mode.
lower_threshold	Direct lower-tail threshold in direct mode.
verbose	Logical. If TRUE, prints a formatted threshold-model summary. Set to FALSE to suppress all console output.

Details

Cases are selected above the upper threshold and controls below the lower threshold. The middle portion is excluded, so the affected and unaffected selection events are not complements. Percentile inputs refer to standard normal population percentiles before conditioning on genotype. The genotype-specific normal mixture follows Eq. 6.1 (pp. 324–325). Genotype-specific upper- and lower-tail probabilities, their population prevalences, and conditional genotype frequencies follow Eqs. 6.3–6.5 (p. 326). Eq. 6.6 (p. 327) is a numerical worked example, not an additional symbolic step. This is a classical/Falconer-style parameterization as formulated for this design framework by Gordon, Finch, and Kim (2020).

Value

An object of class "qtl_falconer_threshold_parameters" containing thresholds, genotype-specific penetrances, selected-population prevalences, conditional genotype frequencies, and Falconer parameters.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.1, Eqs. 6.3–6.5, p. 326. doi:10.1007/9783030611217.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th ed. Longman.

See Also

[qtl_falconer_parameters](#), [qtl_threshold_chisq_power](#), and [qtl_threshold_chisq_mssn](#).

Examples

```
qtl_falconer_threshold_parameters(
  qtl_var = 0.025, tau = 0.5, pd = 0.15,
  x_upper = 5, x_lower = 5
)
```

qtl_multivariate_mssn_full

Multivariate Falconer Minimum Sample Size

Description

Finds the minimum integer total sample size for Pillai MANOVA or the minimum selected case-control sample for a joint-threshold genotype chi-square test.

Usage

```
qtl_multivariate_mssn_full(
  power,
  alpha,
  qtl_var,
  tau,
  pd,
  cor_matrix,
  test = c("pillai", "threshold_chisq"),
  x_upper = NULL,
  x_lower = NULL,
  k = 1,
  verbose = TRUE
)
```

Arguments

power	Target power in (0, 1).
alpha	Significance level in (0, 1).
qtl_var	Numeric vector of phenotype-specific QTL variances in (0, 1).
tau	Numeric vector of phenotype-specific dominance/additivity ratios.
pd	Shared increaser-allele frequency in (0, 1).
cor_matrix	Positive-definite phenotype correlation matrix. Its order must match qtl_var and tau.
test	Either "pillai" or "threshold_chisq".
x_upper, x_lower	Vectors of upper- and lower-tail percentages. A case satisfies every upper threshold (joint AND); a control satisfies every lower threshold (joint AND).
k	Control/case ratio for the threshold chi-square design.
verbose	Logical; whether to print a polished summary.

Details

Pillai MSSN is found by an integer search because its denominator degrees of freedom depend on sample size. For threshold selection, the statistical MSSN is kept separate from the expected population screening burden. The Pillai branch recomputes its noncentral F distribution at each candidate total N; the returned integer is the first design attaining the target power. The historical fractional root is retained only for comparison. The threshold branch uses direct multivariate-normal rectangle integration from Eqs. 6.11–6.14 and genotype chi-square Eq. 1.22; it solves for selected cases and controls and reports source-population screening expectations separately. The two branches therefore return different sample-size units.

Value

An object of class "qtl_multivariate_mssn_full". Both modes include test, target and achieved power, alpha, and the full falconer model. The Pillai result includes integer total N, historical_fractional_mssn,

genotype counts, NCP, degrees of freedom, critical value, and all pillai intermediates. The threshold result includes selected case/control/total MSSNs, target and achieved NCPs, internal S, thresholds and integration diagnostics, expected genotype counts, sparse-cell diagnostics, and separately labelled expected source-population screening counts.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.2, Eqs. 6.9–6.14, pp. 332–333; validation in Section 6.2.4, pp. 336–339. doi:10.1007/9783030611217.
- Gordon, D., Londono, D., Patel, P., Kim, W., Finch, S. J., & Heiman, G. A. (2017). An analytic solution to computation of power and sample size for genetic association studies under a pleiotropic mode of inheritance. *Human Heredity*, 81(4), 194–209. doi:10.1159/000457135.
- Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26(1), 117–121. doi:10.1214/aoms/1177728599.
- Genz, A., & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer. doi:10.1007/9783642016899.

See Also

[qtl_multivariate_power_full](#), [qtl_falconer_parameters](#), and [qtl_threshold_chisq_mssn](#).

Examples

```
cor_matrix <- matrix(c(1, 0.15, 0.15, 1), 2, byrow = TRUE)
qtl_multivariate_mssn_full(
  power = 0.95, alpha = 5e-8, qtl_var = c(0.01, 0.005),
  tau = c(0, 0.5), pd = 0.25, cor_matrix = cor_matrix,
  test = "pillai", verbose = FALSE
)
```

qtl_multivariate_power_full

Multivariate Falconer Power

Description

Computes prospective power for a one-way MANOVA using Pillai's trace or for a genotype chi-square test after joint multivariate threshold selection.

Usage

```
qtl_multivariate_power_full(
  N = NULL,
  N_case = NULL,
  alpha,
```

```

    qtl_var,
    tau,
    pd,
    cor_matrix,
    test = c("pillai", "threshold_chisq"),
    x_upper = NULL,
    x_lower = NULL,
    k = 1,
    verbose = TRUE
)

```

Arguments

N	Integer total sample size for test = "pillai".
N_case	Selected case sample size for test = "threshold_chisq".
alpha	Significance level in (0, 1).
qtl_var	Numeric vector of phenotype-specific QTL variances in (0, 1).
tau	Numeric vector of phenotype-specific dominance/additivity ratios.
pd	Shared increaser-allele frequency in (0, 1).
cor_matrix	Positive-definite phenotype correlation matrix. Its order must match qtl_var and tau.
test	Either "pillai" or "threshold_chisq".
x_upper, x_lower	Vectors of upper- and lower-tail percentages. A case satisfies every upper threshold (joint AND); a control satisfies every lower threshold (joint AND).
k	Control/case ratio for the threshold chi-square design.
verbose	Logical; whether to print a polished summary.

Details

Genotype-specific means are stored in a matrix whose rows are phenotypes and columns are genotypes 0, 1, and 2. All genotypes share the residual covariance matrix formed by scaling `cor_matrix` by residual standard deviations. Joint probabilities use `mvtnorm::pmvnorm()` with the deterministic Miwa algorithm through 20 dimensions; higher dimensions use reproducibly seeded Genz–Bretz integration and return its error/status.

Pillai power follows Gordon et al.'s general matrix derivation: the characteristic roots of Φ -star determine V -star and $\lambda = N s V\text{-star} / (s - V\text{-star})$. The null critical value and alternative power use central and noncentral F distributions.

The two test branches are distinct. With test = "pillai", N is a total sample size and the function performs a three-genotype one-way MANOVA using Pillai's trace. With test = "threshold_chisq", N_{case} is a selected case count; cases must exceed every upper threshold and controls must fall below every lower threshold, after which their conditional genotype probabilities are compared by a two-df chi-square test. The latter is not a MANOVA.

The multivariate mixture and component densities correspond to textbook Eqs. 6.9–6.10; rectangular probabilities to Eq. 6.11; and prevalence and bound construction to Eqs. 6.12–6.14 (Chapter 6, Section 6.2, pp. 332–333). The threshold branch then uses genotype chi-square Eq. 1.22 (p. 26). Eq. 6.15 is not used by this direct-integration implementation.

Value

An object of class "qtl_multivariate_power_full". Both modes contain test, alpha, power, and falconer, the complete multivariate mixture model. For test = "pillai", additional components include total N, genotype frequencies/counts, the NCP, numerator and denominator degrees of freedom, critical value, and pillai, containing contrasts, characteristic roots, Pillai trace, and matrix intermediates. For test = "threshold_chisq", components include selected N_case, N_control, and N_total, k, two-df NCP, internal S, thresholds (bounds, penetrances, prevalences, conditional genotype frequencies, and integration diagnostics), expected genotype counts, and sparse-cell diagnostics.

References

- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Chapter 6, Section 6.2, Eqs. 6.9–6.14, pp. 332–333; validation in Section 6.2.4, pp. 336–339. doi:10.1007/9783030611217.
- Gordon, D., Londono, D., Patel, P., Kim, W., Finch, S. J., & Heiman, G. A. (2017). An analytic solution to computation of power and sample size for genetic association studies under a pleiotropic mode of inheritance. *Human Heredity*, 81(4), 194–209. doi:10.1159/000457135.
- Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26(1), 117–121. doi:10.1214/aoms/1177728599.
- Genz, A., & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer. doi:10.1007/9783642016899.

See Also

[qtl_multivariate_mssn_full](#), [qtl_falconer_parameters](#), and [qtl_threshold_chisq_power](#).

Examples

```
cor_matrix <- matrix(c(1, 0.15, 0.15, 1), 2, byrow = TRUE)
qtl_multivariate_power_full(
  N = 4514, alpha = 5e-8, qtl_var = c(0.01, 0.005),
  tau = c(0, 0.5), pd = 0.25, cor_matrix = cor_matrix,
  test = "pillai", verbose = FALSE
)
```

qtl_threshold_chisq_mssn

Minimum Sample Size for a Threshold-Selected Genotype Chi-Square Test

Description

Calculates the minimum selected case and control sample sizes after deriving their genotype distributions from a single-trait Falconer threshold model.

Usage

```
qtl_threshold_chisq_mssn(
  power,
  alpha,
  qtl_var,
  tau,
  pd,
  x_upper,
  x_lower,
  k = 1,
  verbose = TRUE
)
```

Arguments

power	Target power in $(0, 1)$.
alpha	Significance level in $(0, 1)$.
qtl_var	Variance in the standardized trait attributable to the QTL. Must be in $(0, 1)$.
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in $(0, 1)$.
x_upper	Percentage selected from the upper population tail in percentile mode.
x_lower	Percentage selected from the lower population tail in percentile mode.
k	Positive control-to-case ratio $N_{control}/N_{case}$.
verbose	Logical. If TRUE, prints a concise summary.

Details

The primary MSSN is the selected case-control sample size used by the association test. The separately reported screening quantities are expected population counts needed to obtain those selected samples under simple population sampling; they are not statistical MSSN values. Genotype-specific distributions and threshold selection follow Eqs. 6.1 and 6.3–6.5 (pp. 324–326). The downstream genotype-test NCP is Eq. 1.22 (p. 26). The function solves for selected cases and controls; expected screening counts are reported separately and are not rounded MSSNs.

Value

An object of class "qtl_threshold_chisq_mssn" containing selected MSSN values, target NCP, internal S, thresholds, penetrances, prevalences, conditional frequencies, Falconer parameters, and separately labelled expected population screening counts.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Eqs. 1.22, 6.1, and 6.3–6.5, pp. 26 and 324–326. doi:10.1007/9783030611217.

See Also

[qtl_threshold_chisq_power](#) and [qtl_falconer_threshold_parameters](#).

Examples

```
qtl_threshold_chisq_mssn(
  power = 0.8, alpha = 0.0001,
  qtl_var = 0.025, tau = 0.5, pd = 0.15,
  x_upper = 5, x_lower = 5, verbose = FALSE
)
```

qtl_threshold_chisq_power

Power for a Threshold-Selected Genotype Chi-Square Test

Description

Calculates power for the 2-df genotype chi-square test after a quantitative trait is converted into upper-tail cases and lower-tail controls.

Usage

```
qtl_threshold_chisq_power(
  N_case,
  alpha,
  qtl_var,
  tau,
  pd,
  x_upper,
  x_lower,
  k = 1,
  verbose = TRUE
)
```

Arguments

N_case	Positive number of selected cases.
alpha	Significance level in $(0, 1)$.
qtl_var	Variance in the standardized trait attributable to the QTL. Must be in $(0, 1)$.
tau	Finite dominance-to-additivity ratio δ/a .
pd	Frequency of the increaser allele. Must be in $(0, 1)$.
x_upper	Percentage selected from the upper population tail in percentile mode.
x_lower	Percentage selected from the lower population tail in percentile mode.
k	Positive control-to-case ratio $N_{control}/N_{case}$.
verbose	Logical. If TRUE, prints a concise summary.

Details

Genotype-specific normal means and the common residual variance follow the mixture in Eq. 6.1. Upper-tail cases and lower-tail controls are constructed with Eqs. 6.3–6.5 (pp. 324–326); the middle of the trait distribution is excluded. The resulting conditional genotype probabilities are compared by the two-degree-of-freedom genotype chi-square test using Eq. 1.22 (p. 26). N_{case} and $k * N_{\text{case}}$ are selected case and control counts, not numbers screened from the source population.

Value

An object of class "qtl_threshold_chisq_power" containing selected sample sizes, power, non-centrality parameter, internal effect component S , thresholds, penetrances, prevalences, conditional genotype frequencies, and Falconer parameters.

References

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer, Eqs. 1.22, 6.1, and 6.3–6.5, pp. 26 and 324–326. doi:10.1007/9783030611217.

See Also

[qtl_threshold_chisq_mssn](#) and [qtl_falconer_threshold_parameters](#).

Examples

```
qtl_threshold_chisq_power(
  N_case = 126, alpha = 0.0001,
  qtl_var = 0.025, tau = 0.5, pd = 0.15,
  x_upper = 5, x_lower = 5, verbose = FALSE
)
```

tdt_expected_nontransmission_probability

Expected Non-Transmission Probability (g_{NT}^) in the Transmission Disequilibrium Test (TDT)*

Description

Computes the expected non-transmission probability (g_{NT}^*) for the Transmission Disequilibrium Test (TDT) under a general model including linkage disequilibrium (LD), relative risks, and phenotype misclassification. Implements Equation 5.25 from *Gordon et al. (2020)*, *Heterogeneity in Statistical Genetics*.

Usage

```
tdt_expected_nontransmission_probability(
  pd,
  prev,
  R1,
  R2,
  delta_prime = 1,
  pi01 = 0,
  theta1 = NULL,
  digits = 6,
  verbose = TRUE
)
```

Arguments

pd	Numeric. Frequency of the disease-associated allele.
prev	Numeric. Disease prevalence (ϕ_1).
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
delta_prime	Numeric. Linkage disequilibrium scale parameter (D') (default = 1).
pi01	Numeric. Misclassification rate (default = 0).
theta1	Numeric. Population allele frequency (defaults to pd if NULL).
digits	Integer. Number of digits for printing (default = 6).
verbose	Logical. If TRUE, prints intermediate quantities (default = TRUE).

Details

The function derives intermediate penetrance-weighted frequencies f_0, f_1, f_2 and the contrast term C , which captures differences in genotype contributions. Linkage disequilibrium is represented by $D = D'p_d p_+$. The expected non-transmission probability is then:

$$g_{NT}^* = \frac{p_d p_+ \phi_1 + D(\theta_1 - p_d)C + \pi_{01}(p_d p_+ \phi_0 + D(p_d - \theta_1)C)}{\phi_1 + \pi_{01}\phi_0}$$

When $\pi_{01} = 0$, this reduces to the standard non-misclassified case. This is Eq. 5.25 in Chapter 5, Section 5.2.6 (pp. 284–285). g_{NT_star} is a probability, not an expected count.

Value

A list containing:

gNT_star	Expected non-transmission probability (g_{NT}^*).
C, D	Contrast and LD terms.
f0, f1, f2	Derived penetrance frequencies.
pd, phi1, phi0, theta1	Input and derived parameters.

References

Buyske, S., Yang, G., Matisse, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981. PMID: 19172087.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_expected_transmission_probability](#), [tdt_expected_transmission_counts](#), and [tdt_power](#).

Examples

```
# Example: compute gNT* under no misclassification
tdt_expected_nontransmission_probability(
  pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi01 = 0
)

# Example: with 10% misclassification
tdt_expected_nontransmission_probability(
  pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi01 = 0.1
)
```

tdt_expected_transmission_counts

Expected Transmissions (ET) and Non-Transmissions (ENT*) in the Transmission Disequilibrium Test (TDT)*

Description

Computes the expected number of transmissions (ET^*) and non-transmissions (ENT^*) for a given number of affected trios under heterogeneity and linkage disequilibrium models. Implements Equations 5.31-5.32 from *Gordon et al. (2020), Heterogeneity in Statistical Genetics*.

Usage

```
tdt_expected_transmission_counts(
  N_star,
  pd,
  prev,
  R1,
  R2,
```

```

    delta_prime = 1,
    pi = 1,
    theta1 = NULL,
    digits = 6,
    verbose = TRUE
)

```

Arguments

N_star	Numeric. Number of affected trios.
pd	Numeric. Frequency of the disease-associated allele.
prev	Numeric. Disease prevalence (ϕ_1).
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
delta_prime	Numeric. Linkage disequilibrium scale parameter (D') (default = 1).
pi	Numeric in $[0, 1]$. Linked/homogeneous trio fraction (default 1); $1 - \pi$ is the heterogeneous fraction.
theta1	Numeric. Population allele frequency (defaults to pd if NULL).
digits	Integer. Number of digits for printing (default = 6).
verbose	Logical. If TRUE, prints intermediate quantities (default = TRUE).

Details

The expected transmissions (ET^*) and non-transmissions (ENT^*) are computed as:

$$ET^* = 2N^* \left[\pi \left(\frac{D(p_+ - \theta_1)C}{\phi_1} + p_d p_+ \right) + (1 - \pi) \left(\frac{D(p_+ - 0.5)C}{\phi_1} + p_d p_+ \right) \right]$$

$$ENT^* = 2N^* \left[\pi \left(\frac{D(\theta_1 - p_d)C}{\phi_1} + p_d p_+ \right) + (1 - \pi) \left(\frac{D(0.5 - p_d)C}{\phi_1} + p_d p_+ \right) \right]$$

These are Eqs. 5.31–5.32 in Chapter 5, Section 5.3.3 (pp. 293–294). They are expected counts over N_{star} affected trios. Dividing by $2N^*$ gives the corresponding g_T^* and g_{NT}^* probabilities. The canonical heter_rate used by `tdt_power()` and `tdt_mssn()` equals $1 - \pi$.

Value

A list containing:

ET_star	Expected transmissions (ET^*).
ENT_star	Expected non-transmissions (ENT^*).
C, D	Contrast and LD terms.
f0, f1, f2	Derived penetrance frequencies.
pd, prev, R1, R2, pi, theta1	Input parameters.

References

- Chen, C., Yang, G., Buyske, S., Matise, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501. PMID: 19883370.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_power](#), [tdt_mssn](#), and [tdt_mssn_from_model](#).

Examples

```
# Example 1: Homogeneous model (pi = 1)
tdt_expected_transmission_counts(
  N_star = 1000, pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi = 1
)

# Example 2: Heterogeneous model (pi = 0.7)
tdt_expected_transmission_counts(
  N_star = 1000, pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi = 0.7
)
```

tdt_expected_transmission_probability

Expected Transmission Probability (g_T^) in the Transmission Disequilibrium Test (TDT)*

Description

Computes the expected transmission probability (g_T^*) for the Transmission Disequilibrium Test (TDT), incorporating effects of linkage disequilibrium (LD), relative risks, and phenotype misclassification. Implements Equation 5.24 from *Gordon et al. (2020), Heterogeneity in Statistical Genetics*.

Usage

```
tdt_expected_transmission_probability(
  pd,
  prev,
  R1,
  R2,
  delta_prime = 1,
```

```

    pi01 = 0,
    theta1 = NULL,
    digits = 6,
    verbose = TRUE
  )

```

Arguments

pd	Numeric. Frequency of the disease-associated allele.
prev	Numeric. Disease prevalence (ϕ_1).
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
delta_prime	Numeric. Linkage disequilibrium scale parameter (D') (default = 1).
pi01	Numeric. Misclassification rate (default = 0).
theta1	Numeric. Population allele frequency (defaults to pd if NULL).
digits	Integer. Number of digits for printed results (default = 6).
verbose	Logical. If TRUE, prints intermediate quantities (default = TRUE).

Details

The function computes the penetrance-weighted genotype frequencies f_0, f_1, f_2 and the contrast term C , representing genotype-specific effects. Linkage disequilibrium is modeled as $D = D'p_d p_+$. The expected transmission probability is given by:

$$g_T^* = \frac{p_d p_+ \phi_1 + D(p_+ - \theta_1)C + \pi_{01}(p_d p_+ \phi_0 + D(\theta_1 - p_+)C)}{\phi_1 + \pi_{01}\phi_0}$$

When $\pi_{01} = 0$, this reduces to the standard non-misclassified case. This is Eq. 5.24 in Chapter 5, Section 5.2.6 (p. 284). gT_star is a probability, not an expected count.

Value

A list containing:

gT_star	Expected transmission probability (g_T^*).
C, D	Contrast and LD terms.
f0, f1, f2	Derived penetrance frequencies.
pd, phi1, phi0, theta1	Input and derived parameters.

References

- Buyske, S., Yang, G., Matisse, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981. PMID: 19172087.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_expected_nontransmission_probability](#), [tdt_expected_transmission_counts](#), and [tdt_power](#).

Examples

```
# Example 1 - no misclassification
tdt_expected_transmission_probability(
  pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi01 = 0
)

# Example 2 - 10% misclassification
tdt_expected_transmission_probability(
  pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1, pi01 = 0.1
)
```

tdt_mssn

Family-Based (TDT) Minimum Sample Size Necessary

Description

Computes the minimum number of affected trios required to achieve a specified power for the transmission disequilibrium test (TDT) with no effect or one selected effect: phenotype misclassification, locus heterogeneity, or genotype misclassification. `input_mode` lets the transmission probabilities come either from a genetic model ("model_based", the default) or directly from user-supplied expected transmission and non-transmission counts ("model_free").

Usage

```
tdt_mssn(
  target_power,
  input_mode = c("model_based", "model_free"),
  pd = NULL,
  prev = NULL,
  R1 = NULL,
  R2 = NULL,
  alpha = 0.05,
  delta_prime = 1,
  misclass_rate = 0,
  heter_rate = 0,
  ET = NULL,
  ENT = NULL,
  n_trios = NULL,
  verbose = TRUE,
```

```

    effect = c("none", "phenotype_misclassification", "locus_heterogeneity",
              "genotype_misclassification"),
    genotype_misclassification_rate = NULL
)

```

Arguments

target_power	Numeric in (0,1). Desired power for the TDT.
input_mode	Character. One of "model_based" (default) or "model_free". See Details.
pd	Numeric in (0,1). Frequency of the disease (high-risk) allele at the marker locus. Required for input_mode = "model_based". Optional for "model_free": required only if heter_rate or misclass_rate is non-zero, and if omitted in that case it is solved from ET/ENT (see Details).
prev	Numeric in (0,1). Disease prevalence (ϕ_1). Required for input_mode = "model_based". Required for "model_free" only if misclass_rate is non-zero.
R1	Numeric > 0. Genotype relative risk for heterozygotes. Required for input_mode = "model_based"; unused for "model_free".
R2	Numeric > 0. Genotype relative risk for homozygotes. Required for input_mode = "model_based"; unused for "model_free".
alpha	Numeric in (0,1). Significance level for the TDT (default 0.05).
delta_prime	Numeric. Linkage disequilibrium scale parameter D' (default 1). Used for input_mode = "model_based" only.
misclass_rate	Numeric in [0,1). Phenotype misclassification rate for controls (π_{01}) in the misclassification scenario. Zero means no misclassification and is the default.
heter_rate	Numeric in [0,1). Proportion of trios whose affection status is not due to the locus of interest ($1 - \pi$) in the heterogeneity scenario. Zero means complete homogeneity and is the default.
ET, ENT	Numeric ≥ 0 . Expected transmission and non-transmission counts for input_mode = "model_free".
n_trios	Numeric > 0. Number of affected trios that the supplied ET and ENT correspond to. Required for input_mode = "model_free" (unlike the power function, there is no N argument here to reuse, since this function solves for the required number of trios).
verbose	Logical. If TRUE (default), prints a formatted summary of the no-error required number of trios and any requested modifier-specific required counts and percentage increases.
effect	Character. Optional explicit effect selector: "none", "phenotype_misclassification", "locus_heterogeneity", or "genotype_misclassification". The default is no effect. Calls that omit this argument retain the historical separate-sensitivity behavior of misclass_rate and heter_rate.
genotype_misclassification_rate	Numeric in [0,0.5]. The Chapter 5 adjacent-genotype TDT error parameter e , required only when effect = "genotype_misclassification".

Details

With `input_mode = "model_based"`, penetrances are derived from `prev`, `R1`, `R2`, and `pd`, and

$$N^* = \frac{\lambda^*(g_T^* + g_{NT}^*)}{2(g_T^* - g_{NT}^*)^2}$$

is computed for each of the three scenarios. The no-error calculation follows Eq. 1.25 (p. 27); phenotype misclassification follows Eqs. 5.24–5.27 (pp. 284–285); and locus heterogeneity follows Eqs. 5.30–5.34b (pp. 293–294). Eq. 5.28 is a numerical example. Here `heter_rate` is the heterogeneous fraction, equal to $1 - \pi$ when lower-level formulas use homogeneous fraction π . Phenotype misclassification and locus heterogeneity are reported as separate sensitivity scenarios; supplying both rates does not create a combined-error scenario.

With `input_mode = "model_free"`, the no-error scenario uses $g_T = ET/(2 n_{trios})$ and $g_{NT} = ENT/(2 n_{trios})$. The misclassification and heterogeneity scenarios then use the same closed-form identities as `tdt_power` (see its Details for the formulas and the `pd`-solving fallback), applied to this no-error g_T/g_{NT} pair.

For `effect = "genotype_misclassification"`, the Chapter 5 single-parameter adjacent-genotype TDT transition model, Mendelian-inconsistency filtering, and Eqs. 5.16–5.22 are the same as in `tdt_power`. The adjusted NCP is linear in the number of affected trios, so the existing target-NCP is divided by the adjusted per-trio NCP. This model-based branch cannot be combined with phenotype misclassification or locus heterogeneity. TDTae and sequencing/NGS genotype error are not implemented, and the result uses the package's existing nominal asymptotic TDT convention.

Value

An object of class `"tdt_mssn"`, containing:

alpha, target_power, lambda_star, input_mode Design targets, target one-df NCP, and input mode.

N Named `no_error`, `misclassification`, and `heterogeneity` required affected-trio counts.

percent_increase Modifier-specific conventional percentage inflation, $100(N_{adjusted}/N_{no-error} - 1)$, relative to the no-error required count. Values are NA when the no-error required count is not finite and positive.

power_at_N_no_error, power_loss_at_N_no_error Power and power loss obtained if the no-error design size is used under each modifier.

gT_star, gNT_star Scenario-specific transmission and non-transmission probabilities used in the MSSN formulas.

model_parameters Supplied and derived genetic-model, LD, misclassification, heterogeneity, and model-free count-scale information.

References

- Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *American Journal of Human Genetics*, 52(3), 506–516. PMID: 8447318; PMCID: PMC1682161.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

Buyske, S., Yang, G., Matisse, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981.

Chen, C., Yang, G., Buyske, S., Matisse, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501.

See Also

[tdt_power](#) for the power counterpart.

Examples

```
# Model-based input
tdt_mssn(
  target_power = 0.80, input_mode = "model_based",
  pd = 0.30, prev = 0.05, R1 = 1.5, R2 = 2.25,
  verbose = FALSE
)$N$no_error

# Chapter 5 single-parameter adjacent-genotype TDT error
tdt_mssn(
  target_power = 0.80,
  pd = 0.30, prev = 0.05, R1 = 1.5, R2 = 2.25,
  effect = "genotype_misclassification",
  genotype_misclassification_rate = 0.01,
  verbose = FALSE
)$N$genotype_misclassification

# model_free: supply expected transmissions/non-transmissions directly
tdt_mssn(
  target_power = 0.80, input_mode = "model_free",
  ET = 140, ENT = 100, n_trios = 120,
  pd = 0.30, prev = 0.05,
  verbose = FALSE
)$N$no_error
```

tdt_mssn_from_model	<i>TDT Minimum Sample Size Necessary from Genetic Model Parameters</i>
---------------------	--

Description

Computes the required number of affected trios (N^*) needed to achieve a specified statistical power in the Transmission Disequilibrium Test (TDT), given model parameters for allele frequency, relative risks, disease prevalence, and heterogeneity. Implements the probability and MSSN chain in Eqs. 5.33–5.34b (Chapter 5, Section 5.3.3, pp. 293–294) of Gordon, Finch, and Kim (2020).

Usage

```
tdt_mssn_from_model(
  power,
  alpha,
  df,
  pd,
  prev,
  R1,
  R2,
  delta_prime = 1,
  pi = 1
)
```

Arguments

power	Numeric. Desired power (e.g., 0.8).
alpha	Numeric. Significance level (e.g., 0.05).
df	Integer. Degrees of freedom (typically 1 for TDT).
pd	Numeric. Frequency of the disease-associated allele.
prev	Numeric. Disease prevalence.
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
delta_prime	Numeric. Linkage disequilibrium (LD) scale factor (default = 1).
pi	Numeric in $[0, 1]$. Linked/homogeneous trio fraction; 1 is complete homogeneity and $1 - \pi$ is the heterogeneous fraction.

Details

This function determines the non-centrality parameter (λ^*) via root-finding (`uniroot`) such that the test power equals the desired level. It then computes the expected transmission (gT^*) and non-transmission (gNT^*) probabilities, followed by the required number of trios using:

$$N^* = \frac{\lambda^*}{2} \frac{(gT^* + gNT^*)}{(gT^* - gNT^*)^2}$$

The expected transmission and non-transmission components are calculated under allele frequency and penetrance model assumptions in Eq. 5.33. Their NCP and MSSN are Eqs. 5.34a–b.

Value

A list containing:

lambda_star	Non-centrality parameter (λ^*).
gT_star	Expected transmission probability.
gNT_star	Expected non-transmission probability.
N_star	Required number of trios.

References

Chen, C., Yang, G., Buyske, S., Matise, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_mssn](#), [tdt_power_from_model](#), and [tdt_expected_transmission_counts](#).

Examples

```
# Example: compute required trios for 80% power at alpha = 0.05
tdt_mssn_from_model(
  power = 0.8, alpha = 0.05, df = 1,
  pd = 0.25, prev = 0.005, R1 = 2, R2 = 2,
  delta_prime = 1, pi = 1
)
```

tdt_mssn_phenotype_misclassification_from_ncp

*TDT Minimum Sample Size Necessary from an NCP under Phenotype
Misclassification*

Description

Computes the expected transmissions (gT^*) and non-transmissions (gNT^*) as well as the required number of trios (N^*) for a specified non-centrality parameter (λ^*) under a misclassification model. Implements Eq. 5.26 and Eqs. 5.27a–b (Chapter 5, Section 5.2.6, pp. 284–285) of Gordon, Finch, and Kim (2020). Eq. 5.28a–b is the book’s numerical worked example.

Usage

```
tdt_mssn_phenotype_misclassification_from_ncp(
  lambda_star,
  pd,
  prev,
  R1,
  R2,
  delta_prime = 1,
  pi01 = 0,
  digits = 5
)
```

Arguments

lambda_star	Numeric. Non-centrality parameter (λ^*) derived from desired power (e.g., from <code>tdt_mssn_from_model()</code>).
pd	Numeric. Frequency of the disease-associated allele.
prev	Numeric. Disease prevalence (ϕ_1).
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
delta_prime	Numeric. Linkage disequilibrium scale parameter (D') (default = 1).
pi01	Numeric. Misclassification rate for controls (default = 0).
digits	Integer. Number of digits to round intermediate printed results (default = 5).

Details

This function adjusts the Transmission Disequilibrium Test (TDT) for possible phenotype misclassification. When $\pi_{01} > 0$, a fraction of unaffected individuals are incorrectly classified as affected, inflating the apparent transmission probability. The expected transmission (gT^*) and non-transmission (gNT^*) are calculated as:

$$gT^* = p_d p_+ + \frac{D(p_T - \theta_1)C(1 - \pi_{01})}{\phi_1 + \pi_{01}\phi_0}$$

$$gNT^* = p_d p_+ + \frac{D(p_A - \theta_1)C(\pi_{01} - 1)}{\phi_1 + \pi_{01}\phi_0}$$

The required number of trios is then:

$$N^* = \frac{\lambda^*(gT^* + gNT^*)}{2(gT^* - gNT^*)^2}$$

Setting $\pi_{01} = 0$ reproduces the standard (non-misclassified) TDT.

Value

A list containing:

gT_star	Expected transmission probability.
gNT_star	Expected non-transmission probability.
N_required	Required number of trios (N^*).
lambda_star	Non-centrality parameter.
C, f0, f1, f2	Intermediate derived values used in Eq. 5.26.

References

- Buyske, S., Yang, G., Matisse, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981. PMID: 19172087.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_mssn](#), [tdt_expected_transmission_probability](#), and [tdt_expected_nontransmission_probability](#).

Examples

```
# Example: Compute N* with misclassification adjustment (pi01 = 0.1)
tdt_mssn_phenotype_misclassification_from_ncp(
  lambda_star = 7.8488,
  pd = 0.25, prev = 0.005,
  R1 = 2, R2 = 2,
  delta_prime = 1,
  pi01 = 0.1
)
```

tdt_ngs_mssn

*Analytic MSSN for a TDT1-NGS Sequencing Study***Description**

Computes the prospective minimum sample size necessary (MSSN), measured in complete father-mother-affected-child trios, for the same single-variant TDT1-NGS design as `tdt_ngs_power()`.

Usage

```
tdt_ngs_mssn(
  power,
  pd,
  R1,
  coverage = NULL,
  seq_error = NULL,
  alpha = 0.05,
  verbose = TRUE,
  father_coverage = coverage,
  mother_coverage = coverage,
  child_coverage = coverage,
  epsilon0 = seq_error,
  epsilon1 = seq_error,
  pheno_misclass = FALSE,
  prev = NULL,
  pi01 = 0,
  locus_het = FALSE,
  heter_rate = 0
)
```

Arguments

power	A single finite target power strictly between 0 and 1.
pd	A single finite disease/risk-allele frequency strictly between 0 and 1. The supplied allele labeling is retained.
R1	A single finite positive heterozygote genotype relative risk under the multiplicative model. The homozygote relative risk is $R_2 = R_1^2$.
coverage	A single finite integer greater than or equal to 2. Common fixed-depth shorthand for the father, mother, and affected child. It may be omitted when all three member-specific depths are supplied.
seq_error	A single finite symmetric per-read sequencing-error probability in $[0, 0.5)$. Internally, the directional error parameters default to $\epsilon_0 = \epsilon_1 = \text{seq_error}$. It may be omitted when both directional parameters are supplied.
alpha	A single finite significance level in $(0, 1)$. Defaults to 0.05.
verbose	Logical scalar. If TRUE, print a concise result summary.
father_coverage, mother_coverage, child_coverage	Optional fixed depths for each trio member. An explicitly supplied member depth overrides coverage; otherwise that member inherits coverage.
epsilon0	Directional probability that a reference-allele read is observed as the alternative allele. Defaults to seq_error.
epsilon1	Directional probability that an alternative-allele read is observed as the reference allele. Defaults to seq_error.
pheno_misclass	Logical scalar. If TRUE, add a phenotype- misclassification sensitivity scenario using the ordinary-TDT NCP bridge.
prev	Disease prevalence in $(0, 1)$. Required when either phenotype misclassification or locus heterogeneity is requested. It is used only by the ordinary-TDT bridges and does not alter the raw-read likelihood.
pi01	Probability that an unaffected individual is misclassified or ascertained as affected, in $[0, 1)$. It must be zero unless pheno_misclass = TRUE.
locus_het	Logical scalar. If TRUE, add an independent locus- heterogeneity sensitivity scenario using the ordinary-TDT NCP bridge.
heter_rate	Heterogeneous trio fraction in $[0, 1]$. The ordinary- TDT linked/homogeneous fraction is $\pi = 1 - \text{heter_rate}$. It must be zero unless locus_het = TRUE.

Details

The calculation uses Kim's Appendix B TDT1-NGS noncentrality parameter

$$\lambda = N\delta^2 I_{eff},$$

where N is the number of complete trios, I_{eff} is the nuisance-adjusted per-trio information under the null, and under the multiplicative model $\delta = \log(R_1)$. The target one-degree-of- freedom NCP, λ_* , is obtained by numerical inversion of the noncentral chi-square power function. Trio MSSN is then solved analytically:

$$N_{continuous} = \frac{\lambda_*}{\log(R_1)^2 I_{eff}}.$$

The planned MSSN is the ceiling of this quantity, with achieved power and the immediately smaller design checked at the integer boundary. For the optional phenotype scenario, the ordinary-TDT attenuation factor r_{pheno} described in [tdt_ngs_power](#) multiplies the per-trio coefficient before inversion:

$$N_{continuous,pheno} = \frac{\lambda_*}{\log(R_1)^2 I_{eff} r_{pheno}}.$$

The bridge fixes $R_2 = R_1^2$, $\delta' = 1$, and $\theta_1 = p_d$. It does not modify Kim's likelihood or information matrix and is not a TDTae or simulation model.

For locus heterogeneity, r_{het} is the ratio of the ordinary-TDT heterogeneous and homogeneous per-trio NCP coefficients, with $\pi = 1 - \text{heter_rate}$. The adjusted coefficient is $\log(R_1)^2 I_{eff} r_{het}$ before inversion. Phenotype and heterogeneity requests remain separate scenarios; no product of their attenuation factors is used.

If target power is no greater than alpha, the target NCP is zero and the minimum supported design is one trio. If $R_1 = 1$ and target power exceeds alpha, no finite MSSN exists because the transmission effect is zero.

Each member uses its resolved fixed coverage, and the directional read model is $q_G = \epsilon_0 + (1 - \epsilon_0 - \epsilon_1)G/2$. See [tdt_ngs_power](#) for parameter precedence and interpretation. Coverage 1 is unsupported for every member because efficient information is not identifiable under the implemented 11-parameter nuisance model. This prospective calculation uses raw sequencing read-count information and latent trio genotype states; it performs no simulation, genotype calling, or EM fitting.

Value

Invisibly, an object of class "tdt_ngs_mssn" containing the target, continuous and integer trio requirements, achieved power and NCP, per-trio NCP coefficient, multiplicative-model parameters, efficient information, the 11 by 11 information matrix, numerical diagnostics, effective sequencing-design metadata, model metadata, and scenario-level base and independently adjusted coefficients and sample sizes. Legacy coverage and seq_error fields retain supplied shorthand.

References

- Kim, W. (2015). Transmission disequilibrium tests based on read counts for low-coverage next-generation sequence data. *Human Heredity*, 80(1), 36–49. doi:10.1159/000434645.
- Buyske, S., Yang, G., Matise, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981.
- Chen, C., Yang, G., Buyske, S., Matise, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_ngs_power](#)

Examples

```

tdt_ngs_mssn(
  power = 0.80, pd = 0.325, R1 = 1.2,
  coverage = 12, seq_error = 0.005,
  alpha = 5e-8, verbose = FALSE
)
tdt_ngs_mssn(
  power = 0.80, pd = 0.325, R1 = 1.2, coverage = 12, seq_error = 0.005,
  locus_het = TRUE, prev = 0.01, heter_rate = 0.25,
  alpha = 5e-8, verbose = FALSE
)
tdt_ngs_mssn(
  power = 0.80, pd = 0.325, R1 = 1.2, coverage = 12, seq_error = 0.005,
  pheno_misclass = TRUE, prev = 0.01, pi01 = 0.05,
  alpha = 5e-8, verbose = FALSE
)

```

tdt_ngs_power

Analytic Power for a TDT1-NGS Sequencing Study

Description

Computes prospective analytic power for a single-variant TDT1-NGS study of complete father-mother-affected-child trios. The implementation uses the published latent-state, sequencing-read-count likelihood rather than hard genotype calls.

Usage

```

tdt_ngs_power(
  N,
  pd,
  R1,
  coverage = NULL,
  seq_error = NULL,
  alpha = 0.05,
  verbose = TRUE,
  father_coverage = coverage,
  mother_coverage = coverage,
  child_coverage = coverage,
  epsilon0 = seq_error,
  epsilon1 = seq_error,
  pheno_misclass = FALSE,
  prev = NULL,
  pi01 = 0,
  locus_het = FALSE,
  heter_rate = 0
)

```

Arguments

N	A single finite integer greater than or equal to 1. Number of complete father-mother-affected-child trios.
pd	A single finite disease/risk-allele frequency strictly between 0 and 1. The null information is reflection-symmetric in pd and 1 - pd; the supplied allele labeling is retained.
R1	A single finite positive heterozygote genotype relative risk under the multiplicative model. The homozygote relative risk is $R_2 = R_1^2$.
coverage	A single finite integer greater than or equal to 2. Common fixed-depth shorthand for the father, mother, and affected child. It may be omitted when all three member-specific depths are supplied.
seq_error	A single finite symmetric per-read sequencing-error probability in $[0, 0.5)$. Internally, the directional error parameters default to $\epsilon_0 = \epsilon_1 = \text{seq_error}$. It may be omitted when both directional parameters are supplied.
alpha	A single finite significance level in $(0, 1)$. Defaults to 0.05.
verbose	Logical scalar. If TRUE, print a concise result summary.
father_coverage, mother_coverage, child_coverage	Optional fixed depths for each trio member. An explicitly supplied member depth overrides coverage; otherwise that member inherits coverage.
epsilon0	Directional probability that a reference-allele read is observed as the alternative allele. Defaults to seq_error.
epsilon1	Directional probability that an alternative-allele read is observed as the reference allele. Defaults to seq_error.
pheno_misclass	Logical scalar. If TRUE, add a phenotype- misclassification sensitivity scenario using the ordinary-TDT NCP bridge.
prev	Disease prevalence in $(0, 1)$. Required when either phenotype misclassification or locus heterogeneity is requested. It is used only by the ordinary-TDT bridges and does not alter the raw-read likelihood.
pi01	Probability that an unaffected individual is misclassified or ascertained as affected, in $[0, 1)$. It must be zero unless pheno_misclass = TRUE.
locus_het	Logical scalar. If TRUE, add an independent locus- heterogeneity sensitivity scenario using the ordinary-TDT NCP bridge.
heter_rate	Heterogeneous trio fraction in $[0, 1]$. The ordinary- TDT linked/homogeneous fraction is $\pi = 1 - \text{heter_rate}$. It must be zero unless locus_het = TRUE.

Details

TDT1-NGS is evaluated for one biallelic variant under Hardy-Weinberg parental genotype frequencies, random mating, a multiplicative disease model, fixed member-specific coverage, and common directional sequencing error. With $t = R_1/(1 + R_1)$, the transmission parameter is $\delta = \log\{t/(1 - t)\} = \log(R_1)$ and $R_2 = R_1^2$.

Kim's Appendix B noncentrality parameter is

$$\lambda = N\delta^2 I_{eff},$$

where I_{eff} is the nuisance-adjusted per-trio information evaluated under the null from raw sequencing read-count probabilities and the 15 latent Mendelian trio states. Power is the upper-tail probability beyond the central one-degree-of-freedom chi-square critical value under a noncentral chi-square distribution with NCP λ .

When requested, phenotype misclassification is a sensitivity bridge rather than a phenotype-aware sequencing likelihood. Let $c = 2(g_T - g_{NT})^2 / (g_T + g_{NT})$ be the ordinary-TDT per-trio NCP coefficient. The bridge computes

$$r_{pheno} = c_{pheno} / c_{noerror}$$

from the ordinary-TDT expected transmission and nontransmission probabilities, using the fixed assumptions $R_2 = R_1^2$, $\delta' = 1$, and $\theta_1 = p_d$. It then reports $\lambda_{pheno} = \lambda_{NGS} r_{pheno}$. Kim's raw-read likelihood, information matrix, and efficient information remain unchanged.

Locus heterogeneity uses the analogous independent bridge

$$r_{het} = c_{heterogeneity} / c_{noerror}, \quad \lambda_{het} = \lambda_{NGS} r_{het}.$$

The ordinary-TDT expected counts are evaluated at one trio using $\pi = 1$ for the homogeneous reference and $\pi = 1 - \text{heter_rate}$ for the requested sensitivity scenario. It uses the same fixed R_2 , δ' , and θ_1 assumptions as the phenotype bridge. If both modifiers are requested, they are reported separately against the same sequencing-only baseline; their factors are never multiplied and no combined result is calculated.

The existing directional read model is

$$q_G = \epsilon_0 + (1 - \epsilon_0 - \epsilon_1)G/2,$$

for genotype $G \in \{0, 1, 2\}$. Thus $q_0 = \epsilon_0$, $q_2 = 1 - \epsilon_1$, and $q_1 = (1 + \epsilon_0 - \epsilon_1)/2$; the heterozygote probability is one-half in the symmetric case. The effective errors must be nonnegative and sum to less than one.

Coverage 1 is unsupported under the full published nuisance model. Its eight observable read-count triples provide at most seven independent probability dimensions for an 11-parameter information model, so efficient information is not identifiable without changing the model or using a generalized inverse.

This prospective calculation performs no simulation or EM fitting. It does not implement TDT2-NGS, conventional genotype misclassification/TDTae, a joint modifier likelihood, or multi-locus testing. Coverage is fixed for each member rather than random or sample-specific.

Value

Invisibly, an object of class "tdt_ngs_power" containing the design inputs, power and NCP, multiplicative-model parameters, efficient information, the 11 by 11 information matrix, compact numerical diagnostics, effective member-specific coverage and directional-error metadata, model metadata, and a scenarios list that always contains `sequencing_only` and, when requested, `phenotype_misclassification` and `heterogeneity`. Modifier scenarios record their ordinary-TDT bridge inputs, coefficients, attenuation factor, and confirmation that the raw-read likelihood was not modified. Legacy coverage and `seq_error` fields retain the supplied common shorthand values.

References

- Kim, W. (2015). Transmission disequilibrium tests based on read counts for low-coverage next-generation sequence data. *Human Heredity*, 80(1), 36–49. doi:10.1159/000434645.
- Buyske, S., Yang, G., Matisse, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981.
- Chen, C., Yang, G., Buyske, S., Matisse, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

Examples

```
tdt_ngs_power(
  N = 5000, pd = 0.325, R1 = 1.2,
  coverage = 12, seq_error = 0.005,
  alpha = 5e-8, verbose = FALSE
)
tdt_ngs_power(
  N = 5000, pd = 0.325, R1 = 1.2, coverage = 12, seq_error = 0.005,
  locus_het = TRUE, prev = 0.01, heter_rate = 0.25,
  alpha = 5e-8, verbose = FALSE
)
tdt_ngs_power(
  N = 5000, pd = 0.325, R1 = 1.2, coverage = 12, seq_error = 0.005,
  pheno_misclass = TRUE, prev = 0.01, pi01 = 0.05,
  alpha = 5e-8, verbose = FALSE
)
```

tdt_power

Family-Based (TDT) Power

Description

Computes power for the transmission disequilibrium test (TDT) at a fixed number of affected trios with no effect or one selected effect: phenotype misclassification, locus heterogeneity, or genotype misclassification. `input_mode` lets the transmission probabilities come either from a genetic model ("model_based", the default) or directly from user-supplied expected transmission and non-transmission counts ("model_free").

Usage

```
tdt_power(
  N,
  input_mode = c("model_based", "model_free"),
```

```

pd = NULL,
prev = NULL,
R1 = NULL,
R2 = NULL,
alpha = 0.05,
delta_prime = 1,
misclass_rate = 0,
heter_rate = 0,
ET = NULL,
ENT = NULL,
verbose = TRUE,
effect = c("none", "phenotype_misclassification", "locus_heterogeneity",
  "genotype_misclassification"),
genotype_misclassification_rate = NULL
)

```

Arguments

N	Numeric > 0. Number of affected trios.
input_mode	Character. One of "model_based" (default) or "model_free". See Details.
pd	Numeric in (0,1). Frequency of the disease/high-risk allele at the marker locus. Required for input_mode = "model_based". Optional for "model_free": required only if heter_rate or misclass_rate is non-zero, and if omitted in that case it is solved from ET/ENT (see Details).
prev	Numeric in (0,1). Disease prevalence (ϕ_1). Required for input_mode = "model_based". Required for "model_free" only if misclass_rate is non-zero.
R1	Numeric > 0. Genotype relative risk for heterozygotes. Required for input_mode = "model_based"; unused for "model_free".
R2	Numeric > 0. Genotype relative risk for homozygotes. Required for input_mode = "model_based"; unused for "model_free".
alpha	Numeric in (0,1). Significance level for the TDT (default 0.05).
delta_prime	Numeric. Linkage disequilibrium scale parameter D' (default 1). Used for input_mode = "model_based" only.
misclass_rate	Numeric in [0, 1). Phenotype misclassification rate for controls (π_{01}). A value of 0 corresponds to no misclassification and is the default.
heter_rate	Numeric in [0, 1). Proportion of trios whose affection status is <i>not</i> due to the locus of interest ($1 - \pi$). A value of 0 corresponds to complete homogeneity and is the default.
ET, ENT	Numeric ≥ 0 . Expected transmission and non-transmission counts for input_mode = "model_free", assumed to be accumulated over the same N trios that power is computed for.
verbose	Logical. If TRUE (default), prints a formatted summary of no-error power and any requested modifier-specific power.
effect	Character. Optional explicit effect selector: "none", "phenotype_misclassification", "locus_heterogeneity", or "genotype_misclassification". The default

is no effect. Calls that omit this argument retain the historical separate-sensitivity behavior of `misclass_rate` and `heter_rate`.

`genotype_misclassification_rate`

Numeric in $[0, 0.5]$. The Chapter 5 adjacent-genotype TDT error parameter e , required only when `effect = "genotype_misclassification"`.

Details

With `input_mode = "model_based"`, penetrances f_0, f_1, f_2 are derived from `prev`, `R1`, `R2`, and `pd`, and the expected transmission and non-transmission probabilities under each of the three scenarios are computed from the standard TDT formulation in Eq. 1.25 (Chapter 1, Section 1.6.1.3, p. 27), the phenotype-misclassification probabilities in Eqs. 5.24–5.27 (Section 5.2.6, pp. 284–285), and the locus-heterogeneity construction in Eqs. 5.30–5.34a (Section 5.3.3, pp. 293–294) of Gordon, Finch, and Kim (2020). Eq. 5.28 is a numerical worked example, not the general symbolic formula. Phenotype misclassification and locus heterogeneity are reported as separate sensitivity scenarios; supplying both rates does not create a combined-error scenario.

N is the number of affected-child trios, with both parents genotyped. ET and ENT are expected counts accumulated over trios, whereas g_T and g_{NT} are per-parental-allele transmission and non-transmission probabilities. `heter_rate` is the heterogeneous fraction; in lower-level formulas written with the linked/homogeneous fraction π ; equivalently, the heterogeneous fraction is $1 - \pi$.

With `input_mode = "model_free"`, the no-error scenario uses $g_T = ET/(2N)$ and $g_{NT} = ENT/(2N)$ directly. The misclassification and heterogeneity scenarios are then computed from the following identities, which are algebraically equivalent to the the corresponding model-based identities. Let $A = g_T - g_{NT}$ (no-error) and $p_+ = 1 - p_d$:

heterogeneity: $g_T = p_d p_+ + A(p_+ - 0.5 \times \text{heter_rate})$, $g_{NT} = p_d p_+ + A(-p_d + 0.5 \times \text{heter_rate})$,

$$\text{misclassification: } m = \frac{\phi_1(1 - \pi_{01})}{\phi_1 + \pi_{01}(1 - \phi_1)}, \quad g_T = p_d p_+ + p_+ A m, \quad g_{NT} = p_d p_+ - p_d A m.$$

A scenario is only computed from these identities when its rate is non-zero; at a rate of exactly 0 the scenario reuses the no-error g_T/g_{NT} directly, so `pd` is not required unless at least one rate is non-zero, and `prev` is not required unless `misclass_rate` is non-zero.

If `pd` is needed but not supplied, it is solved from ET and ENT via $2p_d^2 - 2p_d(1 - A) + (g_T + g_{NT} - A) = 0$. This quadratic has two roots summing to $1 - A$; the root in $(0, 0.5)$ is used, with a message reporting the derived value. An error is raised if no such unique root exists – supplying `pd` directly is preferred.

The genotype-misclassification effect implements the Chapter 5 single-parameter adjacent-genotype TDT model in Section 5.2.5: $\theta_{02} = \theta_{20} = 0$ and $\theta_{01} = \theta_{10} = \theta_{12} = \theta_{21} = e$. The independent member transitions are applied to the 15 true Mendelian-consistent trios, producing 27 observed configurations (Eqs. 5.18–5.20). Observed Mendelian-inconsistent trios are removed and the retained probabilities are renormalized (Eqs. 5.16–5.17); expected transmitted and non-transmitted counts and the NCP then follow Eqs. 5.22 and 5.21. This branch requires model-based input and cannot be combined with phenotype misclassification or locus heterogeneity. Only this Chapter 5 single-parameter adjacent-genotype TDT model is supported: TDTae and sequencing/NGS genotype error are not implemented. Power uses the package's existing nominal asymptotic chi-square threshold; conventional genotype error can inflate the ordinary TDT Type-I error.

Value

An object of class "tdt_power", containing:

alpha, N, input_mode Significance level, affected-trio count, and input mode.

lambda, power Named no_error, misclassification, and heterogeneity NCP and power values.

power_loss Absolute power loss for each modifier relative to the no-error scenario.

gT_star, gNT_star Scenario-specific transmission and non-transmission probabilities.

ET, ENT Scenario-specific expected transmission and non-transmission counts for the supplied N trios.

model_parameters Supplied and derived genetic-model, LD, misclassification, and heterogeneity parameters.

References

Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *American Journal of Human Genetics*, 52(3), 506–516. PMID: 8447318; PMCID: PMC1682161.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

Buyske, S., Yang, G., Matise, T. C., & Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4), 287–292. doi:10.1159/000194981.

Chen, C., Yang, G., Buyske, S., Matise, T., Finch, S. J., & Gordon, D. (2009). Transmission disequilibrium test power and sample size in the presence of locus heterogeneity. *Statistical Applications in Genetics and Molecular Biology*, 8, Article 44. doi:10.2202/15446115.1501.

See Also

[tdt_mssn](#) for the MSSN counterpart.

Examples

```
# Model-based input
tdt_power(
  N = 600, input_mode = "model_based",
  pd = 0.30, prev = 0.05, R1 = 1.5, R2 = 2.25,
  misclass_rate = 0.01, heter_rate = 0.10,
  verbose = FALSE
)$power$no_error

# Chapter 5 single-parameter adjacent-genotype TDT error
tdt_power(
  N = 600, pd = 0.30, prev = 0.05, R1 = 1.5, R2 = 2.25,
  effect = "genotype_misclassification",
  genotype_misclassification_rate = 0.01,
  verbose = FALSE
)$power$genotype_misclassification
```

```
# model_free: supply expected transmissions/non-transmissions directly
tdt_power(
  N = 600, input_mode = "model_free",
  ET = 140, ENT = 100,
  pd = 0.30, prev = 0.05,
  misclass_rate = 0.01, heter_rate = 0.10,
  verbose = FALSE
)$power$no_error
```

tdt_power_from_expected_counts

Transmission Disequilibrium Test (TDT) Power from Expected Transmissions and Non-Transmissions

Description

Computes the statistical power of the Transmission Disequilibrium Test (TDT) given the expected number of transmissions (ET) and non-transmissions (ENT) under a specified significance level. Implements Eq. 1.25 (Chapter 1, Section 1.6.1.3, p. 27) of Gordon, Finch, and Kim (2020).

Usage

```
tdt_power_from_expected_counts(ET, ENT, alpha = 0.05)
```

Arguments

ET	Numeric. Expected number of transmissions.
ENT	Numeric. Expected number of non-transmissions.
alpha	Numeric. Significance level (default = 0.05).

Details

The function calculates the non-centrality parameter and statistical power for the TDT using the chi-square distribution with 1 degree of freedom. ET and ENT are expected counts, not probabilities, and must refer to the same affected-trio design.

The non-centrality parameter is computed as:

$$\lambda = \frac{(ET - ENT)^2}{ET + ENT}$$

Power is then obtained as:

$$1 - P(\chi_{1,\lambda}^2 < q_{\chi_{1,1-\alpha}^2})$$

Value

A list containing:

lambda	Non-centrality parameter.
power	Computed power at the given alpha level.
ET	Expected transmissions.
ENT	Expected non-transmissions.

References

Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *American Journal of Human Genetics*, 52(3), 506–516. PMID: 8447318; PMCID: PMC1682161.

Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_power](#) and [tdt_power_from_model](#).

Examples

```
# Example: compute power for ET = 140 and ENT = 100
tdt_power_from_expected_counts(ET = 140, ENT = 100, alpha = 0.05)
```

tdt_power_from_model	<i>Transmission Disequilibrium Test (TDT) Power from Genetic Model Parameters</i>
----------------------	---

Description

Computes the statistical power of the Transmission Disequilibrium Test (TDT) using genetic model parameters such as allele frequency, relative risks, disease prevalence, and the number of affected trios. Implements the penetrance construction in Eqs. 1.6–1.7 (p. 13) and TDT Eq. 1.25 (p. 27) of Gordon, Finch, and Kim (2020).

Usage

```
tdt_power_from_model(
  pd,
  N,
  delta_prime,
  f0 = NULL,
  f1 = NULL,
  f2 = NULL,
  prev = NULL,
```

```

    R1 = NULL,
    R2 = NULL,
    alpha = 0.05
)

```

Arguments

pd	Numeric. Frequency of the disease-associated allele.
N	Numeric. Number of affected trios.
delta_prime	Numeric. Linkage disequilibrium (LD) scale factor (default = 1).
f0, f1, f2	Optional. Penetrances for genotypes with 0, 1, and 2 risk alleles. If not provided, they are computed internally using prev, R1, and R2.
prev	Numeric. Disease prevalence.
R1	Numeric. Relative risk for heterozygotes.
R2	Numeric. Relative risk for homozygotes.
alpha	Numeric. Significance level (default = 0.05).

Details

When penetrances (f0, f1, f2) are not provided, they are derived from the model parameters using:

$$f0 = prev/Z, \quad f1 = R1 * f0, \quad f2 = R2 * f0$$

where

$$Z = (1 - p_d)^2 + 2 * p_d * (1 - p_d) * R1 + p_d^2 * R2.$$

Expected transmission (*ET*) and non-transmission (*ENT*) counts are computed based on the allele frequency and penetrance model, and the power is derived from the non-central chi-square distribution with 1 degree of freedom. Here N is the number of affected-child trios. delta_prime scales $D = p_d(1 - p_d)D'$ in the implemented model; the disease-locus and marker-locus assumptions should be considered when interpreting pd.

Value

A list containing:

lambda	Non-centrality parameter.
power	Computed power at the given significance level.
ET	Expected transmissions.
ENT	Expected non-transmissions.
Penetrances	Vector of computed penetrances (f0, f1, f2).

References

- Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *American Journal of Human Genetics*, 52(3), 506–516. PMID: 8447318; PMCID: PMC1682161.
- Gordon, D., Finch, S. J., & Kim, W. (2020). *Heterogeneity in Statistical Genetics: How to Assess, Address, and Account for Mixtures in Association Studies*. Springer. doi:10.1007/9783030611217.

See Also

[tdt_power](#), [tdt_power_from_expected_counts](#), and [tdt_mssn_from_model](#).

Examples

```
# Example: model-based power computation
tdt_power_from_model(
  pd = 0.25, N = 10000, delta_prime = 1,
  prev = 0.05, R1 = 1, R2 = 1.1, alpha = 0.05
)
```

Index

case_control_genotype_misclassification, [3](#), [10](#), [12](#), [17](#), [28](#)
 case_control_locus_heterogeneity, [7](#), [8](#), [17](#), [28](#)
 case_control_phenotype_misclassification, [7](#), [11](#), [17](#), [28](#)
 cc_chisq_mssn_genotype_misclassification_1p (case_control_genotype_misclassification), [3](#)
 cc_chisq_mssn_genotype_misclassification_2p (case_control_genotype_misclassification), [3](#)
 cc_chisq_mssn_genotype_misclassification_3p (case_control_genotype_misclassification), [3](#)
 cc_chisq_mssn_genotype_misclassification_differential_3p (case_control_genotype_misclassification), [3](#)
 cc_chisq_mssn_locus_heterogeneity (case_control_locus_heterogeneity), [8](#)
 cc_chisq_mssn_phenotype_misclassification (case_control_phenotype_misclassification), [11](#)
 cc_chisq_power_genotype_misclassification_1p (case_control_genotype_misclassification), [3](#)
 cc_chisq_power_genotype_misclassification_2p (case_control_genotype_misclassification), [3](#)
 cc_chisq_power_genotype_misclassification_3p (case_control_genotype_misclassification), [3](#)
 cc_chisq_power_genotype_misclassification_differential_3p (case_control_genotype_misclassification), [3](#)
 cc_chisq_power_locus_heterogeneity (case_control_locus_heterogeneity), [8](#)
 cc_chisq_power_phenotype_misclassification (case_control_phenotype_misclassification), [11](#)
 cc_mssn, [3](#), [7](#), [10–12](#), [13](#), [20](#), [28](#), [37](#), [38](#)
 cc_ngs_mssn, [17](#), [41](#)
 cc_ngs_power, [18](#), [20](#), [21](#), [42](#), [43](#)
 cc_power, [3](#), [7](#), [10–12](#), [17](#), [23](#), [24](#), [38](#), [40](#)
 cc_trend_mssn_locus_heterogeneity (case_control_locus_heterogeneity), [8](#)
 cc_trend_power_locus_heterogeneity (case_control_locus_heterogeneity), [8](#)
 lrtae_mssn, [29](#)
 lrtae_ncp, [31](#)
 lrtae_ncp(), [31](#), [33](#), [34](#)
 lrtae_power, [33](#)
 lrtae_power(), [31](#)
 ngs_genotype_error_matrix, [35](#)
 paweh_app, [36](#)
 plot_cc_mssn, [37](#), [40](#)
 plot_cc_power, [38](#), [38](#)
 plot_ngs_mssn, [40](#), [43](#)
 plot_ngs_power, [41](#), [42](#)
 plot_qtl_anova_mssn, [43](#), [45](#)
 plot_qtl_anova_power, [44](#), [44](#)
 plot_qtl_anova_power(), [47](#)
 plot_qtl_genotype_distribution, [45](#)
 plot_qtl_threshold_chisq_mssn, [47](#), [49](#)
 plot_qtl_threshold_chisq_power, [48](#), [49](#)
 plot_qtl_threshold_chisq_power(), [47](#)
 plot_tdt_mssn, [50](#), [53](#), [54](#), [56](#)
 plot_tdt_mssn(), [61](#)
 plot_tdt_mssn_locus_heterogeneity, [51](#), [52](#), [54](#)
 plot_tdt_mssn_phenotype_misclassification, [51](#), [53](#), [53](#)

plot_tdt_power, [51](#), [55](#), [57](#), [59](#)
plot_tdt_power(), [61](#)
plot_tdt_power_locus_heterogeneity, [56](#),
[56](#), [59](#)
plot_tdt_power_phenotype_misclassification,
[56](#), [57](#), [58](#)
plot_tdt_surface3d, [59](#)

qtl_anova_mssn, [43](#), [44](#), [62](#), [64](#), [66](#)
qtl_anova_power, [44](#), [45](#), [63](#), [63](#), [66](#)
qtl_falconer_parameters, [63](#), [64](#), [65](#), [67](#),
[69](#), [71](#)
qtl_falconer_parameters(), [47](#)
qtl_falconer_threshold_parameters, [66](#),
[66](#), [73](#), [74](#)
qtl_multivariate_mssn_full, [67](#), [71](#)
qtl_multivariate_power_full, [69](#), [69](#)
qtl_threshold_chisq_mssn, [47](#), [48](#), [67](#), [69](#),
[71](#), [74](#)
qtl_threshold_chisq_power, [49](#), [67](#), [71](#), [73](#),
[73](#)

shiny::runApp(), [36](#)

tdt_expected_nontransmission_probability,
[74](#), [80](#), [87](#)
tdt_expected_transmission_counts, [76](#),
[76](#), [80](#), [85](#)
tdt_expected_transmission_probability,
[76](#), [78](#), [87](#)
tdt_mssn, [50](#), [51](#), [53](#), [54](#), [78](#), [80](#), [85](#), [87](#), [96](#)
tdt_mssn(), [59](#), [61](#)
tdt_mssn_from_model, [78](#), [83](#), [100](#)
tdt_mssn_phenotype_misclassification_from_ncp,
[85](#)
tdt_ngs_mssn, [41](#), [87](#)
tdt_ngs_power, [42](#), [43](#), [89](#), [90](#)
tdt_power, [55–57](#), [59](#), [76](#), [78](#), [80](#), [82](#), [83](#), [93](#),
[98](#), [100](#)
tdt_power(), [59](#), [61](#)
tdt_power_from_expected_counts, [97](#), [100](#)
tdt_power_from_model, [85](#), [98](#), [98](#)