ST2132 Cheatsheet
by Wei En & Yiyang, AY21/22

ST2131 Topics

Theorems & Identities

Tail Sum Formula

For DRV. X with *non-negative integer-valued* support, $E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$.
For CRV. X with positive support, $E(X) = \int_0^{\infty} P(X > x)\,dx = \int_0^{\infty} P(X \geq x)\,dx$

Markov's Inequality

For **non-negative** r.v. X, $P(X \geq a) \leq \frac{E(X)}{a}$ for any $a > 0$.

Chebyshev's Inequality

Let X be a r.v. with mean μ, $P(|X-\mu| \geq a) \leq \frac{\text{var}(X)}{a^2}$ for any $a > 0$.

One-sided Chebyshev's Inequality

Let X be a r.v. with **zero mean** and variance σ^2, $P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$ for any $a > 0$.

Jensen's Inequality

For r.v. X and convex function $g(X)$, $E[g(X)] \geq g(E[X])$, provided the expectations exist and are finite.

Definitions

Covariance, $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$
Coefficient of Correlation, $\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$
Moment Generating Function, $M_X(t) = E[e^{tX}]$

DRV

Bernoulli

$X \sim Be(p)$, indicate whether an event is successful.

$$P(X=k) = p^k(1-p)^{(1-k)},\ k = 0 \text{ or } 1$$

Statistics: $E(X) = p$, $\text{var}(X) = pq = p(1-p)$
MGF: $M_X(t) = 1 - p + pe^t$

Binomial

$X \sim Bin(n,p)$, total number of successes in n i.i.d. $Be(p)$ trials.

$$P(X=k) = \binom{n}{k} p^x q^{n-x},\ k = 0, 1, \ldots, n$$

Statistics: $E(X) = np$, $\text{var}(X) = npq = np(1-p)$
MGF: $M_X(t) = (1 - p + pe^t)^n$

Geometric

$X \sim Geom(p)$, where $X = 1, 2, \ldots$. Memoryless Property.

$$P(X=k) = pq^{k-1},\ k = 1, 2, \ldots$$

Statistics: $E(X) = \frac{1}{p}$, $\text{var}(X) = \frac{1-p}{p^2}$

MGF: $M_X(t) = \frac{pe^t}{1-qe^t}$

Negative Binomial

$X \sim NB(r,p)$, where $X = r, r+1, \ldots$

$$P(X=k) = \binom{k-1}{r-1} p^r q^{x-r},\ k = r, r+1, \ldots$$

Statistics: $E(X) = \frac{r}{p}$, $\text{var}(X) = \frac{r(1-p)}{p^2}$

Poisson

$X \sim Poisson(\lambda)$

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, \ldots$$

Statistics: $E(X) = \text{var}(X) = \lambda$
MGF: $M_X(t) = e^{\lambda(e^t - 1)}$
Properties: $Poisson(\alpha) + Poisson(\beta) = Poisson(\alpha + \beta)$

Hypergeometric

Suppose there are N identical balls, m of them are red and $N-m$ are blue. $X \sim H(n, N, m)$ is #red balls in n draws without replacement.

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}},\ k = 0, 1, \ldots, n$$

Statistics: $E(X) = \frac{nm}{N}$, $\text{var}(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$

CRV

Uniform

$X \sim U(a,b)$

$$f(x) = \frac{1}{b-a},\quad a < x < b$$

Statistics: $E(X) = \frac{a+b}{2}$, $\text{var}(X) = \frac{(b-a)^2}{12}$
MGF: $M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$, $t \neq 0$

Exponential

$X \sim Exp(\lambda)$ for $\lambda > 0$. Memoryless Property

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$
$$F(x) = 1 - e^{-\lambda x}, x \geq 0$$

Statistics: $E(X) = \frac{1}{\lambda}$, $\text{var}(X) = \frac{1}{\lambda^2}$
MGF: $M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$

Normal

$X \sim N(\mu, \sigma^2)$. Special case : $Z \sim N(0,1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)},\ x \in \mathbb{R}$$

Statistics: $E(X) = \mu$, $\text{var}(X) = \sigma^2$
MGF: $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

Gamma

$X \sim Gamma(\alpha, \lambda)$ for shape α, and rate $\lambda > 0$. ($1/\lambda$ is scale parameter)

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\quad x \geq 0$$

Statistics: $E(X) = \frac{\alpha}{\lambda}$, $\text{var}(X) = \frac{\alpha}{\lambda^2}$
MGF: $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha},\quad t < \beta$
Special case: $Exp(\lambda) = Gamma(1, \lambda)$, $\chi^2_n = Gamma(\frac{n}{2}, \frac{1}{2})$
Properties: $Gamma(a,\lambda) + Gamma(b,\lambda) = Gamma(a+b,\lambda)$, and $cX \sim Gamma(\alpha, \frac{\lambda}{c})$
Gamma function $\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1}\,dy$
$\Gamma(1) = 1$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ and $\Gamma(n) = (n-1)!$, $n \in \mathbb{Z}^+$

Beta

$X \sim B(a,b)$ where $a > 0, b > 0$ has support $[0,1]$

$$f(X) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, 0 \leq x \leq 1$$

Statistics: $E(X) = \frac{1}{1+\beta/\alpha}$, $\text{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ Special case: $\text{Unif}(0,1) = B(1,1)$
Beta function $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Chapter 02 - Random Variables

Functions of a Random Variable

Properties of CDF

(*Ch 2.3 Prop. C*) Let $Z = F(X)$, then $Z \sim \text{Unif}(0,1)$

(*Ch 2.3 Prop. D*) Let $U \sim \text{Unif}(0,1)$, and let $X = F^{-1}(U)$, then the CDF of X is F.

Inverse CDF Method

For a r.v. X with CDF F to be generated, let $U = F(X)$ and write it as $X = F^{-1}(U)$, then generate with following steps:

1. Generate u from a $\text{Unif}(0,1)$.
2. Deliver $x = F^{-1}(u)$.

Distribution of a Function of R.V.

For r.v. X with pdf. $f_X(x)$, assume $g(x)$ is a function of X that is **strictly monotonic** and **differentiable**. Then the pdf. of $Y = g(X)$,

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|,\ y = g(x) \text{ for some } x$$

Chapter 03 - Joint Distributions

Joint Distributions

Copula

A **copula**, $C(u,v)$, is a joint CDF where the marginal distributions are standard uniform. It has properties as shown below:

- $C(u,v)$ is defined over $[0,1]\times[0,1]$ and is non-decreasing
- $P(U \le u) = C(u,1)$ and $P(V \le v) = C(1,v)$
- joint density function $c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v) \ge 0$

Construct joint distributions from marginal distributions given using copula: For any two CRVs X and Y and a copula $C(u,v)$ given,

$$F_{XY}(x,y) = C(F_X(x), F_Y(y))$$

is a joint distribution that has marginal distributions $F_X(x)$ and $F_Y(y)$. Correspondingly, the joint density is

$$f_{XY}(x,y) = c(F_X(x), F_Y(y))\, f_X(x) f_Y(y)$$

Farlie Morgenstern Family

For any two CRVs X and Y with their CDFs $F(x)$ and $G(y)$ given, it is shown that for any constant $|\alpha| \le 1$,

$$H(x,y) = F(x)\,G(y)\left[1 + \alpha(1-F(x))(1-G(y))\right]$$

is a bivariate joint CDF of X and Y, with its marginal CDFs equal to $F(x)$ and $G(y)$.
Farlie Morgenstern copula: $C(u,v) = uv(1+\alpha(1-u)(1-v))$ is the copula used in the Farlie Morgenstern Family.

Bivariate Normal Distribution

If X and Y are jointly distributed with bivariate normal,

$$f(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}$$

where $-1 < \rho < 1$ is the correlation coefficient and the other 4 parameters are reflected in marginal distributions,

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2),\ Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

For a joint distribution to be considered bivariate normal, it must satisfy both:

1. Its two marginal distributions are normal
2. The contours for its joint density function are elliptical

Conditional Joint Distributions

Rejection Method

For a r.v. X with density function $f(x)$ to be generated, if $f(x) > 0$ for $a \le x \le b$, then

1. Let $M = M(x)$ s.t. $M(x) \ge f(x)$ for $a \le x \le b$
2. Let $m = m(x) = \frac{M(x)}{\int_a^b M(t)dt}$ (i.e. m is a pdf of support $[a,b]$)
3. Generate T with density m.
4. Generate U which follows $\text{Unif}(0,1)$ independent of T.
5. If $M(T) \times U \le f(T)$ then deliver T; otherwise, go base to Step 1 and repeat.

Functions of Joint Distributions

(*Ch 3.6.2 Prop. A*) Suppose X and Y are jointly distributed and $u = g_1(x,y),\ v = g_2(x,y)$ can be inverted as $x = h_1(u,v),\ y = h_2(u,v)$ then

$$f_{UV}(u,v) = f_{XY}(h_1(u,v), h_2(u,v))\ |J^{-1}(h_1,h_2)|$$

Sum/Quotient of Random Variables

Suppose X and Y are independent and have JDF f. Then for $U = X+Y$,

$$f_U(u) = \int_{-\infty}^{\infty} f(x, u-x)\,dx,$$

and for $V = X/Y$,

$$f_V(v) = \int_{-\infty}^{\infty} |x| f(x, xv)\,dx.$$

Order Statistics

(*Ch 3.7 Thm. A*) Density function of $X_{(k)}$, the k-th order statistics,

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x)[1-F(x)]^{n-k}$$

Chapter 04 - Expected Values

Model for Measurement Error

Let x_0 denotes the true value of a quantity being measured. Then the measurement, X, can be modeled as:

$$X = x_0 + \beta + \epsilon$$

where β is **bias**, a constant and ϵ is the random component of error. $E(\epsilon) = 0$ and $\text{var}(\epsilon) = \sigma^2$.
Mean Squared Error (MSE) is a measure of the overall measurement error,

$$\begin{aligned} \text{MSE} &= E[(X-x_0)^2]\ \text{(Definition)} \\ &= \sigma^2 + \beta^2 \end{aligned}$$

Conditional Expectation & Prediction

Find Expectation & Variance by Conditioning

$$E(Y) = E[E(Y|X)],\ \text{var}(Y) = \text{var}[E(Y|X)] + E[\text{var}(Y|X)]$$

Random Sum

$$E(T) = E(N)E(X),\ \text{var}(T) = [E(X)]^2\text{var} + E(N)\text{var}(X)$$

Predictions

Suppose X and Y are jointly distributed. If X is observed, the predictor of Y that minimises MSE would be

$$h(Y) = E(Y|X)$$

Delta Method

Consider $Y = g(X)$ where the PDF of X is unknown but μ_X and σ_X^2 is known. Then

$$E(Y) \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_x), \quad \text{var}(Y) \approx \sigma_X^2[g'(\mu_X)]^2.$$

Chapter 05 - Limit Theorems

The RV X **converges in probability** to μ if for any $\epsilon > 0$,

$$P(|X-\mu| > \epsilon) \to 0.$$

The RVs $X_1, X_2, \ldots$ with CDFs $F_1, F_2, \ldots$ **converge in distribution** to X with CDF F if

$$\lim_{n\to\infty} F_n(x) = F(X)$$

at every point which F is continuous.

Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of independent RVs. Then $\overline{X_n} = n^{-1}\sum_{i=1}^n X_i$ converges to μ in probability as $n \to \infty$.

Strong Law of Large Numbers

$$P(\lim_{n\to\infty} \overline{X_n} = \mu) = 1.$$

Continuity Theorem

Let F_n be a sequence of CDFs with corresponding MGFs M_n. If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points of F.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent RVs with mean μ and variance σ^2, and CDF F and MGF M defined in a neighbourhood of zero. Let $S_n = \sum_{i=1}^n (X_i - \mu)$. Then

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x), \quad -\infty < x < \infty.$$

Common Convergences in Distribution

(*Tut.5 Qn3*) $Bin(n,p) \xrightarrow{d} Poission(np)$ as $n\to\infty,\ p\to 0$.

(*Tut.5 Qn4*) For X standardised $Gamma(\alpha,\lambda)$, $X \xrightarrow{d} Z$ as $\alpha\to\infty$.

Miscellaneous

(*Tut.5 Qn13*) For sequence $a_n \to a$, $(1+\frac{a_n}{n})^n \to e^a$.

Chapter 06 - Distributions Derived from the Normal Distribution

Common Distributions

Chi-Square Distribution

For independent $Z_1, \ldots, Z_n \sim N(0,1)$,

$$V = \sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

$M_V(t) = (1-2t)^{-n/2}$.

t-distribution

If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ are independent, then

$$T = \frac{Z}{\sqrt{U/n}} \sim t_n.$$

F-distribution

For independent $U \sim \chi^2_m$ and $V \sim \chi^2_n$,

$$W = \frac{U/m}{V/n} \sim F_{m,n}.$$

Related Identities

(*Tut.6 Qn2*) For $X \sim F_{n,m}$, $X^{-1} \sim F_{m,n}$.
(*Tut.6 Qn3*) For $X \sim t_n$, $X^2 \sim F_{1,n}$.

Sample Mean & Variance

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}.$$

Related Identities

For i.i.d $X_1, ..., X_n$ from $N(\mu, \sigma^2)$,
(*Ch 6.2 Thm.A*) $\bar{X}$ and the vector $\langle X_1 - \bar{X}, ..., X_n - \bar{X}\rangle$ are independent.
(*Ch 6.2 Thm.B*) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
(*Ch 6.2 Coro.B*) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

Chapter 08 - Estimation of Parameters and Fitting of Probability Distributions

Let $\hat{\theta}_n$ be an estimate of a parameter θ based on a sample of size n. Then $\hat{\theta}_n$ is **consistent in probability** if it converges in probability to θ as $n \to \infty$.

Method of Moments

1. Calculate low order moments in terms of their parameters.
2. Find expressions for the parameters in terms of the moments.
3. Insert sample moments into the expressions.

Method of Maximum Likelihood

Consider RVs $X_1, \dots, X_n$ with joint PDF $f(x_1, \dots, x_n \mid \theta)$. The **likelihood** of θ is

$$\text{lik}(\theta) = f(x_1, \dots, x_n \mid \theta).$$

If $X_1, \dots, X_n$ are independent, then the **log likelihood** can be expressed as

$$l(\theta) = \sum_{i=1}^{n} \log[f(x_i \mid \theta)].$$

Invariance Property

Let $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ be a mle of $\theta = (\theta_1, \dots, \theta_k)$ in the density $f(x \mid \theta_1, \dots, \theta_k)$. If $\tau(\theta) = (\tau_1(\theta), \dots, \tau_r(\theta))$, $1 \le r \le k$ is a transformation of the parameter space Θ, then a mle of $\tau(\theta)$ is $\tau(\hat{\theta}) = (\tau_1(\hat{\theta}), \dots, \tau_r(\hat{\theta}))$.

Fisher Information

$$I(\theta) = E\left\{\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right]^2\right\}.$$

Large Sample Theory

Under (varying) smoothness conditions on f,

1. The mle from i.i.d. sample is consistent.
2. $$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right].$$
3. The distribution of $\hat{\theta}$ tends towards
$$N\left(\theta_0, \frac{I}{nI(\theta_0)}\right).$$
4. An approximate $100(1-\alpha)\%$ confidence interval for θ_0 is
$$\hat{\theta} \pm z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta})}}.$$

Bayesian Inference

Unknown parameter θ is treated as a distribution, not a value. **Prior distribution** $f_\Theta(\theta)$ represents our knowledge (assumption) about θ before observing data X. After observation, we have a better estimation using the **posterior distribution** $f_{\Theta|X}(\theta|x)$, where

$$f_{\Theta|X} = \frac{f_{X\Theta}(x,\theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)f_\Theta(\theta)}{\int f_{X|\Theta}(x|t)f_\Theta(t)dt}$$

In short, it means

$$\text{Posterior density} \propto \text{likelihood} \times \text{prior density}$$

Large Sample Theory for Bayesian

As $n \to \infty$,

$$\Theta|X \sim N(\hat{\theta}, -[l''(\hat{\theta})]^{-1})$$

where $\hat{\theta}$ is the mle of θ_0.

Bootstrapping Method

For Estimating Sampling Distribution

1. Assume some distribution provides a good fit to the data.
2. Simulate N random samples of size n from the distribution using the estimated parameter $\hat{\theta}$.
3. For each random sample, calculate estimates of the distribution parameters using either Method of Moments or Maximum Likelihood, θ^*_j for $j = 1, ..., N$.
4. Use the N values of estimates θ^*_j to approximate the sampling distributions of the parameters.

For Estimating Confidence Interval

1. Approximate the distribution of $\hat{\theta} - \theta_0$ with that of $\theta^* - \hat{\theta}$.
2. Obtain the lower and upper bounds $\underline{\delta}$ and $\overline{\delta}$
$$P(\theta^* - \hat{\theta} < \underline{\delta}) = P(\theta^* - \hat{\theta} > \overline{\delta}) = \frac{\alpha}{2}$$
3. The CI for θ_0 can then be constructed:
$$P(\hat{\theta} - \overline{\delta} \le \theta_0 \le \hat{\theta} - \underline{\delta}) = 1 - \alpha$$

Estimator Properties

An estimator $\hat{\theta}$ of θ_0 is **consistent** if $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$.

The **efficiency** of $\hat{\theta}$ relative to $\tilde{\theta}$ is

$$\text{eff}(\hat{\theta}, \tilde{\theta}) = \frac{\text{var}(\tilde{\theta})}{\text{var}(\hat{\theta})}.$$

A statistic $T(X_1, \dots, X_N)$ is **sufficient** for θ if the conditional distribution of $X_1, \dots, X_n$ given $T = t$ does not depend on θ for any value of t.

MSE

$\text{MSE} = \text{var}(\hat{\theta}) + (E(\hat{\theta}) - \theta_0)^2$.
If $\hat{\theta}$ is an unbiased estimator of θ_0, $\text{MSE} = \text{var}(\hat{\theta})$.

Cramer-Rao Inequality

Let $T = t(X_1, \dots, X_n)$ be an unbiased estimator of θ. Then, under smoothness assumptions on $f(x \mid \theta)$, $\text{var}(T) \ge 1/nI(\theta)$.

Factorisation Theorem for Sufficiency

$T(X_1, \dots, X_n)$ is sufficient for a parameter of θ iff the joint pdf factors in the form

$$f(x_1, \dots, x_n \mid \theta) = g[T(x_1, \dots, x_n), \theta]h(x_1, \dots, x_n).$$

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an estimator of θ with $E(\hat{\theta}^2) < \infty$ for all θ. Suppose that T is sufficient for θ and let $\tilde{\theta} = E(\hat{\theta} \mid T)$. Then for all θ,

$$E[(\tilde{\theta} - \theta)^2] \le E[(\hat{\theta} - \theta)^2].$$

Testing Hypotheses and Assessing Goodness of Fit

Testing Hypotheses

The **likelihood ratio** is defined as

$$\frac{P(x \mid H_0)}{P(x \mid H_1)} = \frac{P(H_1)}{P(H_0)}\frac{P(H_0 \mid x)}{P(H_1 \mid x)}$$

with H_0 rejected if the likelihood ratio is less than $c = \frac{P(H_1)}{P(H_0)}$.

Uniformly Most Powerful

If H_0 is simple and H_1 is composite, a test that is most powerful for every simple alternative in H_1 is said to be **uniformly most powerful**.
Conditions: 1) H_1 must be one-sided. 2) The threshold for rejection region must be independent of μ_1 of H_1.

Neyman-Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c has significance level α. Then any other test for which the significance level is at most α has power at most that of the likelihood ratio test.

Generalised Likelihood Ratio

The test statistic corresponding to the **generalised likelihood ratio** is

$$\Lambda = \frac{\max_{\theta\in\omega_0}[\mathrm{lik}(\theta)]}{\max_{\theta\in\Omega}[\mathrm{lik}(\theta)]}$$

where ω_0 is the set of all possible values of θ specified by H_0 and similarly for ω_1, with $\Omega = \omega_0 \cup \omega_1$. The threshold λ_0 is chosen such that $P(\Lambda \le \lambda_0 \mid H_0) = \alpha$, the desired significance level.
Under large sample theory,

$$-2\log\Lambda \dot{\sim} \chi^2_\nu$$

where $\nu = \dim\Omega - \dim\omega_0$.

Multinomial Case

Let $O_i = n\hat{p}_i$ and $E_i = np_i(\hat{\theta})$ denotes the observed and estimated cases,

$$-2\log\Lambda = 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right)$$

Pearson's Chi-square Statistic

$$X^2 = \sum_{i=1}^{m} \frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})} \sim \chi^2_\nu,$$

where ν is the number of degrees of freedom. In practice, $np_i(\hat{\theta}) \ge 5$ is required for the approximation to be good.

Poisson Dispersion Test

Given $x_1, \ldots, x_n$ and testing H_0 that the counts are Poisson with a common parameter λ versus H_1 that they are Poisson but with different rates, the likelihood ratio test statistic is

$$-2\log\Lambda = 2\sum_{i=1}^{n} x_i \log\left(\frac{x_i}{\bar{x}}\right) \approx \frac{1}{\bar{x}}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Goodness of Fit

Hanging Diagrams

- historgram: Plot of $n_j - \hat{n}_j$
- rootogram: Plot of $\sqrt{n_j} - \sqrt{\hat{n}_j}$, var-stabilised
- chi-gram: Plot of $\frac{n_j - \hat{n}_j}{\sqrt{\hat{n}_j}}$, var-stabilised

Variance-stabilising transform: a transformation $Y = g(X)$ that makes $\mathrm{var}(Y)$ (approximately) constant using Delta Method.

Probability Plots

- P-P Plot: Plot $F(X_{(k)})$ against $\frac{k}{n+1}$
- Q-Q Plot: Plot $X_{(k)}$ against $F^{-1}(\frac{k}{n+1})$

Tests for Normality

The **coefficient of skewness** is defined as

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3},$$

where the test rejects for large values of $|b_1|$.
The **coefficient of kurtosis** is defined as

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$$

where the test rejects for large values of $|b_2|$.