1 Some Basic Probability Theory and Calculus

Marginals, conditionals

Suppose $(X,Y) \sim f_{XY}(x,y) \implies$ marginals are $f_X(x) = \int f_{XY}(x,y)\,dy$ and $f_Y(y) = \int f_{XY}(x,y)\,dx$. Conditionals are $f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$

Inverse function theorem:
$X \sim f_X(x)$ and $Y = g(X)$. Then if $X \in \mathbb{R} \implies f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| = f_X\left(g^{-1}(y)\right)\left|\frac{1}{g'(g^{-1}(y))}\right|$.
If $X \in \mathbb{R}^n \implies f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\det J_{g^{-1}}(y)\right| = f_X\left(g^{-1}(y)\right)\left|\det J_g\left(g^{-1}(y)\right)\right|^{-1}$

Moment generating function: $X \sim f_X(x) \implies M_X(t) = \mathbb{E}(\exp(t'x))$.
Properties:

1. If $Y = AX + b \implies M_Y(t) = \exp(t'b)\, M_X(A't)$
2. If X, Y independent $\implies M_{X+Y}(t) = M_X(t)\, M_Y(t)$
3. $M_X'(0) = \mathbb{E}(X), M_X''(0) = \mathbb{E}\left(X^2\right), ..., M_X^{(r)}(0) = \mathbb{E}(X^r)$
4. If $\{X_n\}$ is a sequence of r.v and $M_{X_n}(t) \to M_X(t) \implies X_n \to^d X$

Order Statistics:
$X_1, ..., X_n \sim_{i.i.d} f_X(x)$ with cdf $F_X(x)$. And $X_{(k)}$ the $k-th$ order statistic (from smaller to bigger). Then

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} F_X(x)^{k-1}\left(1 - F_X(x)\right)^{n-k} f(x)$$

Special cases: $X_{(1)} \implies f_{X_{(1)}}(x) = n\left(1 - F_X(x)\right)^{n-1} f_X(x)$ and for $X_{(n)} \implies f_{X_{(n)}}(x) = n\left(F_X(x)\right)^{n-1} f_X(x)$

Taylor Expansion: $f(x) - f(x_0) \simeq \sum_{i=1}^{i=k} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}$ with $f^{(k)}(x_0) = \frac{\partial^k f(x)}{\partial x^k}$. If f is multivariate and expand up to $k = 2 \implies f(x) - f(x_0) \simeq J_f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0)^T$ with J_f the Jacobian and H_f the hessian

Properties of exponential

- $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$
- If $a_n \to a \implies \lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a$

2 Most important Distributions

Distribution	**Support**	**pdf**	**cdf**	$\mathbb{E}(X)$	$\mathbb{V}(X)$	$M_X(t)$
Bernoulli (p)	$\{0,1\}$	$p^x(1-p)^x$	–	p	$p(1-p)$	pe^t+1-p
$B(n,p)$	$\{0,1,...,n\}$	$\begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}$	–	np	$np(1-p)$	$(pe^t+1-p)^n$
Poisson (λ)	$\mathbb{N}$	$e^{-\lambda}\frac{\lambda^x}{x!}$	–	λ	λ	$\exp(\lambda(e^t-1))$
$\exp(\lambda)$	$\mathbb{R}_+$	$\frac{1}{\lambda}e^{-\frac{1}{\lambda}x}$ if $x>0$	$1-e^{-\frac{1}{\lambda}x}$	λ	λ^2	$(1-t\lambda)^{-1}$
$\mathbf{\Gamma}(\alpha,\beta)$	$\mathbb{R}_+$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$ if $x>0$	–	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$
$\chi^2(k)$	$\mathbb{R}_+$	$\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}$ if $x>0$	–	k	$2k$	$(1-2t)^{-\frac{k}{2}}$
$N(\mu,\sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$	–	μ	σ^2	$\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right)$
$N(\mu,\Sigma)$	$\mathbb{R}^n$	$\frac{1}{(2\pi)^{\frac{N}{2}}\det(\Sigma)^{\frac{1}{2}}}e^{\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)}$	–	μ	Σ	$\exp\left(\mu't+\frac{1}{2}t'\Sigma t\right)$
$t(p)$	$\mathbb{R}$	$\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}\frac{1}{(p\pi)^{\frac{1}{2}}\left(1+x^2\frac{1}{p}\right)^{\frac{p+1}{2}}}$	–	0	$\frac{p}{p-2}$	–
$F(d_1,d_2)$	$\mathbb{R}_+$	$\frac{\sqrt{\frac{(d_1x)^{d_1}d_2^{d_2}}{(d_1x+d_2)^{d_1+d_2}}}}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$ with $x>0$	–	$\frac{d_2}{d_2-2}$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$	$\nexists$
$\mathbf{B}(a,\beta)$	$[0,1]$	$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	–	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	–

2.1 Some Special Properties of Distributions

Binomial : If $X\sim B(n,p)$, $Y\sim B(m,p)$ and X indep. of $Y \implies X+Y\sim B(m+n,p)$

Poisson: If $X_i\sim$ **Poisson** (λ_i) independent, then $\sum_{i=1}^{i=n}X_i\sim$ **Poisson** $\left(\sum_{i=1}^{i=n}\lambda_i\right)$

Exponential

- If $X\sim\exp(\lambda)\implies X\sim\Gamma(1,\lambda)$
- If $X_i\sim\exp(\lambda)$ independent $\implies \sum_{i=1}^{i=n}X_i\sim\Gamma(n,\lambda)$
- if $X\sim\exp(\lambda)$ and $\alpha>0\implies \alpha X\sim\exp(\alpha\lambda)$

Chi-Squared

- If $X_i \sim N(0,1)$ independent, then $\sum_{i=1}^{i=n} X_i^2 = \chi^2(n)$
- If $Y_i \sim \chi^2_{k_i}$ independent, then $\sum_{i=1}^{i=n} Y_i = \chi^2\left(\sum_{i=1}^{i=n} k_i\right)$
- If $X \sim \chi^2_n \implies X \sim \Gamma\left(\frac{n}{2}, 2\right)$

Gamma

- Gamma function: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$. It satisfies:
    1. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
    2. $\Gamma(n) = n!$ if $n \in \mathbb{N}$
    3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- If $X_i \sim \Gamma(\alpha_i, \theta)$ indep. $\implies \sum X_i \sim \Gamma\left(\sum \alpha_i, \theta\right)$
- If $X \sim \Gamma(\alpha, \theta)$ and $\phi > 0 \implies \phi X \sim \Gamma(\alpha, \phi\theta)$

Normal Distribution

- If $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ independent, then $\sum_i X_i \sim N\left(\sum_i \mu_i, \sum_i \sigma_i^2\right)$
- If $X \sim N\left(\mu, \sigma^2\right)$ and $a, b \in \mathbb{R} \implies aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$
- If $X \sim N(\mu, \Sigma) \in \mathbb{R}^n \implies$ marginals $X_i \sim N(\mu_i, \Sigma_{ii})$
- If $X \sim N(\mu, \Sigma) \implies$ any subvector X_k is multivariate normal
- If $X \sim N(\mu, \Sigma) \implies AX + b \sim N(A\mu + b, A\Sigma A')$

t-student

- Defined as $t = \frac{N(0,1)}{\sqrt{\chi^2(n)/n}} \sim t(n)$
- As $n \to \infty \implies t(n) \to N(0,1)$

F-distribution (Snedecor)

- Defined as $F(d_1, d_2) = \frac{\chi^2(d_1)/d_1}{\chi^2(d_2)/d_2}$

Beta ($\mathbf{B}(\alpha, \beta)$)

- **Beta function** $=$ $B(\alpha, \beta) \equiv \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}$
- If $X \sim \Gamma(\alpha_X, \theta)$ and $Y \sim \Gamma(\alpha_Y, \theta) \implies \frac{X}{X+Y} \sim \mathbf{B}(\alpha_X, \alpha_Y)$
- If $X \sim U[0,1] \implies X^2 \sim \mathbf{B}\left(\frac{1}{2}, 1\right)$

3 Probability Limits

Almost sure convergence: $X_n \to^{a.s} X$ almost surely $\iff \Pr(\omega : X_n(\omega) \to X(\omega)) = 1$

Convergence in Probability: $p\lim_{n\to\infty} X_n = X \iff$ for all $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| < \varepsilon) = 1$

Convergence in Distribution: $X_n \to^{d} X \iff \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for x continuity point of F_X

Convergence in Quadratic Mean: $X_n \to^{c.m} X \iff \mathbb{E}(X_n - X)^2 \to 0$

Implications: Almost sure convergence $\implies$ Convergence in Probability $\implies$ Convergence in Distribution

Convergence in Quadratic mean $\implies$ Convergence in Probability

Slutsky's Theorem: Let $\{X_n\}, \{Y_n\}$ be seq. of r.v

1. If $X_n \to^p X$, $Y_n \to^p Y \implies X_n Y_n \to^p XY$
2. If $X_n \to^p X$, $Y_n \to^p Y \implies X_n + Y_n \to^p X + Y$
3. If $X_n \to^p X$ and $g(x)$ is a continuous function $\implies g(X_n) \to^p g(X)$
4. If $X_n \to^{a.s} X$ and $g(x)$ is a continuous function $\implies g(X_n) \to^{a.s} g(X)$
5. If $X_n \to^d X$ and $g(x)$ is a continuous function $\implies g(X_n) \to^d g(X)$
6. If $X_n \to^p X$ and $Y_n \to^d Y \implies X_n Y_n \to^d XY$

Chebychev's Inequality: Let X be a random variable and $g(X)$ a nonnegative function. Then for all $r > 0 \implies \Pr(g(X) \geq r) \leq \frac{1}{r}\mathbb{E}(g(X))$

Markov's Theorem Let $\{X_n\}$ be an i.i.d random sequence with $\mathbb{E}(X) = \mu < \infty$ and $\mathbb{V}(X) = \sigma^2 < \infty$. Then

$$\begin{aligned} \overline{X}_n &= \frac{1}{n}\sum_{i=1}^{i=n} X_i \to^p X \text{ (weak law of large numbers) and} \\ \overline{X}_n &= \frac{1}{n}\sum_{i=1}^{i=n} X_i \to^{a.s} X \text{ (strong law of large numbers)} \end{aligned}$$

Central Limit Theorem: Let $\{X_n\}$ be an i.i.d random sequence with $\mathbb{E}(X) = \mu < \infty$ and $\mathbb{V}(X) = \Sigma$. Then

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \to^d N(0, \Sigma)$$

Delta method:

- **Univariate:** Suppose $\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \to^d N\left(0, \sigma^2\right)$ and $g(x)$ is a differentiable function. Then

  1. If $g'(\theta) \neq 0 \implies \sqrt{n}\left(g\left(\widehat{\theta}_n\right) - g(\theta)\right) \to^d N\left(0, \left(g'(\theta)\right)^2 \sigma^2\right)$

2. If $g'(\theta)=0$ and $g \in C^2 \implies n\left(g\left(\widehat{\theta}_n\right)-g(\theta)\right)^2 \to^d \sigma^2 \frac{g''(\mu)}{2}\chi_1^2$

- **Multivariate:** Let $\widehat{\theta}_n \in \mathbb{R}^k$ be a sequence of r.v. and $g:\mathbb{R}^k \to \mathbb{R}^m$ a differentiable function at $x=\theta$. Then if $\sqrt{n}\left(\widehat{\theta}_n-\theta\right) \to^d N(0,\Sigma)$ and $J_g(\mu) \neq 0 \implies \sqrt{n}\left(g\left(\widehat{\theta}_n\right)-g(\theta)\right) \to^d N\left(0, J_g(\mu)\Sigma J_g(\mu)'\right)$ with $J_g(\mu)$ the Jacobian of g at μ

4 Basic Properties of Statistics

Unbiasedness: $\widehat{\theta}(\mathbf{X})$ is unbiased $\iff \mathbb{E}_\theta\left(\widehat{\theta}(\mathbf{X})\right)=\theta$

Consistency: $\widehat{\theta}(\mathbf{X})$ is consistent $\iff \widehat{\theta}(\mathbf{X}) \to^p \theta$

Consistency in MSE (minimum squared error): $\widehat{\theta}(\mathbf{X})$ is consistent in MSE $\iff MSE\left(\widehat{\theta}\right) \equiv \mathbb{E}\left(\widehat{\theta}(\mathbf{X})-\theta\right)^2 \equiv \mathbb{V}\left(\widehat{\theta}(\mathbf{X})\right)+\left(\mathbb{E}\left(\widehat{\theta}(\mathbf{X})-\theta\right)\right)^2 \to 0$ as $n\to\infty$

Properties:

1. $\widehat{\theta}(\mathbf{X})$ is consistent in MSE $\iff \mathbb{E}\left(\widehat{\theta}(\mathbf{X})\right)\to\theta$ and $\mathbb{V}\left(\widehat{\theta}(\mathbf{X})\right)\to 0$.
2. If $\widehat{\theta}(\mathbf{X})$ is consistent in MSE $\implies$ is consistent.
3. $\overline{X}$ is consistent in MSE for $\mathbb{E}(X)$

Asymptotic Normality: $\widehat{\theta}(\mathbf{X})$ is assympt. normal $\iff \sqrt{n}\left(\widehat{\theta}(\mathbf{X})-\theta\right) \to^d N\left(0, V_{\widehat{\theta}}\right)$. Asymptotic efficiency $= V_{\widehat{\theta}}^{-1}$

4.1 Some results

Sufficiency and Factorization Theorem: Let $X_i \sim_{i.i.d} f(x \mid \theta)$ and $T(\mathbf{X})$ a statistic on sample $\mathbf{X}=(X_1,X_2,...,X_n)$. $T(\mathbf{X})$ is sufficient if and only if there exist functions g,h such that

$$f(\mathbf{X}\mid\theta)=g(T(\mathbf{X}),\theta)\,h(\mathbf{X})$$

with $f(\mathbf{X}\mid\theta)$ the joint distribution of the sample.

Minimal Sufficiency: $T(\mathbf{X})$ is minimal sufficient $\iff$ is sufficient and, for two samples $\mathbf{X}$ and $\mathbf{Y}$

$$\frac{f(\mathbf{X}\mid\theta)}{f(\mathbf{Y}\mid\theta)} \text{ is not a function of } \theta \iff T(\mathbf{X})=T(\mathbf{Y})$$

Rao-Blackwell Theorem: Let $W(\mathbf{X})$ be an unbiased estimator of θ, and $T(\mathbf{X})$ be a sufficient statistic for θ. Then the statistic $\phi(\mathbf{X})=\mathbb{E}(W\mid T(\mathbf{X}))$ is unbiased and is uniformly better than $W(\mathbf{X})$, in the sense that $\mathbb{V}_\theta(W(\mathbf{X})) \geq \mathbb{V}_\theta(\phi(\mathbf{X}))$ for all $\theta\in\Theta$

Cramer-Rao Inequality: Let $W(\mathbf{X})$ be an unbiased estimator for θ that satisfies $\frac{d}{d\theta}\mathbb{E}_\theta(W(\mathbf{X})) = \int_X \frac{\partial}{\partial\theta}(W(\mathbf{x}) f(\mathbf{x} \mid \theta))\, d\mathbf{x}$. and has finite variance for all θ. Then

$$\mathbb{V}_\theta(W(\mathbf{X})) \geq \frac{1}{n\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f(X \mid \theta)\right)^2\right)} = \frac{1}{J_n(\theta)}$$

With $J_n(\theta)^{-1}$ being the **Cramer-Rao bound**

4.2 Properties of Mean and Variance

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i$ and $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}\left(X_i - \overline{X}\right)^2$ are unbiased estimators of $\mathbb{E}(X)$ and $\mathbb{V}(X)$ respectively
- Both are asymptotically normal
- If $X_i \sim N\left(\mu, \sigma^2\right)$ then
  1. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
  2. $\frac{(n-1)}{\sigma^2} S_X^2 \sim \chi_{n-1}^2$
  3. $\overline{X}$ and S_X^2 are independent
  4. $\frac{\overline{X}-\mu}{\sqrt{S_X^2/n}} \sim t(n-1)$

4.3 Method of Moments Estimator

Let $X_i \sim_{i.i.d} f(X \mid \theta)$ with $\theta \in \mathbb{R}^K$

- Calculate $\mathbb{E}(X^r) = f_r(\theta)$ for all $r = 1, 2, ..., K$
- Replace $\mathbb{E}(X^r)$ with $\overline{X^r}$ (i.e. $\overline{X^r} = f_r(\theta)$)
- Solve for θ the system of R equations in R unknowns.

If $f^{-1} = (f_1(\theta), f_2(\theta), ..., f_K(\theta))^{-1}$ exists and is continuous, then $\widehat{\theta}_{MM}$ is consistent. If f^{-1} is differentiable, then because of delta method it is asymptotically normal

5 Properties of Maximum Likelihood Estimation

Invariance: If $\widehat{\theta}_{MLE}$ is the MLE of θ, then $\tau\left(\widehat{\theta}_{MLE}\right)$ is the MLE of $\tau(\theta)$ for any function τ

Regularity Conditions

1. $X_1, X_2, ..., X_n$ *i.i.d* with $X_i \sim f(x \mid \theta)$
2. **Identifiably:** If $\theta \neq \theta' \implies f(x \mid \theta) \neq f(x \mid \theta')$
3. $f(x \mid \theta)$ has support that does not depend on θ
4. True parameter θ_0 is interior to Θ
5. $\frac{\partial^3 f(x|\theta)}{\partial \theta^3}(\theta)$ exists, is continuous, and satisfies that for all θ_0
6. There exist a function $M_{\theta_0}(x)$ such that $\mathbb{E}_\theta(M_{\theta_0}(X)) < \infty$ and c_{θ_0} such that for all x and for all $\theta \in (\theta_0 - c_{\theta_0}, \theta_0 + c_{\theta_0})$ we have that $\left|\frac{\partial^3 \ln f(x|\theta)}{\partial \theta^3}(\theta)\right| \leq M_{\theta_0}(x)$

Information equality: Under regularity conditions, $\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial \theta} \ln f(X \mid \theta)\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln(X|\theta)}{\partial \theta^2}\right) \equiv I_0(\theta)$, so $J_n(\theta) = nI_1(\theta)$

Asymptotic Properties of MLE

1. $\widehat{\theta}_{MLE}$ is consistent for θ
2. **Asymptotic Normality:** $\sqrt{n}\left(\widehat{\theta}_{MLE} - \theta\right) \to^d N\left(0, I_1(\theta)^{-1}\right)$
3. **Asymptotic Efficiency**: Asymptotically attains cramer-rao bound

6 Bayesian Statistics

Bayes Rule: Given prior $\pi(\theta)$ with support Θ and conditional distribution $f(\mathbf{x} \mid \theta)$, posterior is calculated as

$$\pi(\theta \mid \mathbf{x}) = \frac{\pi(\theta) f(\mathbf{x} \mid \theta)}{\int_\Theta \pi(\theta') f(\mathbf{x} \mid \theta') \, d\theta'} \propto \pi(\theta) f(\mathbf{x} \mid \theta)$$

7 Testing

Type 1 Error: reject $H_0 \mid H_0$ true

Type 2 Error: not reject $H_0 \mid H_0$ false

Power of a test: $\Pr(\text{reject } H_0 \mid H_0 \text{ false})$.

Power function: $P(\theta) = \Pr(\text{reject } H_0 \mid \theta)$

Neyman-Pearson Lemma: Consider test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$. Then there exist a UMP test, with rejection region

$$\text{reject} \iff \frac{f(\mathbf{X} \mid \theta_1)}{f(\mathbf{X} \mid \theta_0)} > k$$

Mean Test:

1. $H_0 : \mu = \mu_0$ vs. $\mu > \mu_0$. Use test statistic

$$t = \sqrt{n}\left(\frac{\overline{X} - \mu_0}{S_X}\right)$$

and reject $\iff t > K$. If $X \sim N\left(\mu, \sigma^2\right) \implies$ reject if $t > t_{1-\alpha}(n-1)$ being the $1-\alpha$ quantile of $t-$student. If X is not normal, then reject if $t > z_{1-\alpha}$, with z quantile of $N(0,1)$

2. $H_0 : \mu = \mu_0$ vs. $\mu \neq \mu_0$ use same statistic, and If $X \sim N\left(\mu, \sigma^2\right)$ then reject if $t \notin \left(-t_{1-\frac{\alpha}{2}}(n-1), t_{1-\frac{\alpha}{2}}(n-1)\right)$. If not normal, reject if $t \notin \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$

Variance test: Suppose $X \sim N\left(\mu, \sigma^2\right)$ and want to test $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$. Use test statistic

$$\chi^2 = \frac{(n-1)}{\sigma_0^2} S_X^2 \sim \chi^2_{n-1} \text{ under null.}$$

and reject if $\chi^2 < \chi^2_{n-1}(1-\alpha)$. If test $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ use same statistic and reject if $\chi^2 \notin \left[\chi^2_{n-1}\left(\frac{\alpha}{2}\right), \chi^2_{n-1}\left(1 - \frac{\alpha}{2}\right)\right]$

Some critical values:

$$\begin{aligned} z_{0.9} &= 1.2816 & z_{0.975} &= 1.9600 & z_{0.995} &= 2.5758 \\ z_{0.95} &= 1.6449 & z_{0.99} &= 2.3263 \end{aligned}$$

and for χ^2_1 :

$$\begin{aligned} \chi^2_1(0.95) &= 3.8415 & \chi^2_1(0.99) &= 6.6349 & \chi^2_1(0.01) &= 0.00015 \\ \chi^2_1(0.975) &= 5.0239 & \chi^2_1(0.05) &= 0.0039 \end{aligned}$$

7.1 MLE Tests

Want to test $H_0 : \theta = \theta_0$ vs. $\theta \neq \theta_0$

Wald test:

- **Univariate:** $\sqrt{n}\frac{\widehat{\theta}_{MLE} - \theta_0}{\sqrt{I^{-1}(\theta_0)}} \sim N(0,1)$. Do a test like mean test. Or equivalently $n\left(\widehat{\theta}_{MLE} - \theta_0\right)^2 I\left(\widehat{\theta}_{MLE}\right) \simeq \chi^2_1$
- **Multivariate:** $n\left(\widehat{\theta}_{MLE} - \theta_0\right)' I\left(\widehat{\theta}\right)\left(\widehat{\theta}_{MLE} - \theta_0\right) \sim \chi^2_p$ with $p = \#$ of parameters

LR Test

- Let $\lambda(\mathbf{X}) = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}\left(\widehat{\theta}_{MLE}\right)} = \frac{f(\mathbf{X}|\theta_0)}{\max_{\theta \in \Theta} f(\mathbf{X}|\theta)}$. Then $-2\ln(\lambda(\mathbf{X})) \simeq \chi^2_1$ and rejection region is $-2\ln(\lambda(\mathbf{X})) > \chi^2_1(1-\alpha)$.
- In general, if we test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$ use LR $\lambda(\mathbf{X}) = \frac{\max_{\theta \in \Theta_0} f(\mathbf{X}|\theta_0)}{\max_{\theta \in \Theta} f(\mathbf{X}|\theta_0)} = \frac{\mathcal{L}\left(\widehat{\theta}_{\text{RESTRICTED}}\right)}{\mathcal{L}\left(\widehat{\theta}_{MLE}\right)}$ and then $-2\ln(\lambda(\mathbf{X})) \simeq \chi^2_p$ with $p = \#$ restrictions.

LM test (score test)

- **Univariate:** $\frac{1}{\sqrt{n}} \frac{s(\theta_0)}{\sqrt{I(\theta_0)}} \simeq N(0,1)$ and do mean test. Equivalently $\frac{1}{n} s(\theta_0)^2 \frac{1}{I(\theta_0)} \simeq \chi_1^2$
- **Multivariate:** $\frac{1}{n} s(\theta_0)^T I(\theta_0) s(\theta_0) \sim \chi_p^2$ with $p = \#$ of parameters

8 Confidence Sets

Pratt Theorem: Let X be a r.v in $\mathbb{R}$ with $X \sim f(x \mid \theta)$and $C(x) = [L(x), R(x)]$ with both functions increasing. Then for any θ^*

$$\mathbb{E}_{\theta^*}(\text{Length } C(X)) \equiv \mathbb{E}_{\theta}(R(X) - L(X)) = \int_{\theta \neq \theta^*} \Pr(\theta \in C(X) \mid \theta^*)\, d\theta$$

Confidence sets for mean and variance $\implies$ Same as two-sided tests of previous section.