Probability Notes

Mark DiBattista

March 18, 2019

Abstract

Probability.

1 Suggested Resource Materials

Useful source texts:

- Probability/Statistics, intermediate (probability sections are better than statistics): *Statistical Inference*, Casella & Berger
- Probability, advanced: *Probability and Measure*, Billingsley

Throughout the text the acronym, *LAN*, refers to the companion writeup, *Linear Algebra Notes*, in which information is referenced by chapter and/or numbered equation.

2 Probability Preliminaries

2.1 Standard Nomenclature

$$\left.\begin{array}{ll} E: & \text{Event} \\ X, Y: & \text{Random Variables} \end{array}\right\} \Rightarrow \begin{array}{ll} \text{Probability:} & \mathbb{P}E \\ \text{Mean:} & \mathbb{E}X \\ \text{Variance:} & \mathbb{V}X \\ \text{Covariance:} & \mathbb{C}(X,Y) \\ \text{Information:} & \mathbb{I}X \\ \text{Entropy:} & \mathbb{H}X \end{array} \tag{1}$$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \Rightarrow \begin{cases} \mathbb{E}\mathbf{X} & = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{pmatrix} \\ \mathbb{V}\mathbf{X} & = \mathbb{E}(\mathbf{X} - \mathbb{E}\mathbf{X})\,\mathbb{E}(\mathbf{X} - \mathbb{E}\mathbf{X})^\top = \mathbb{E}\mathbf{X}\mathbf{X}^\top - \mathbb{E}\mathbf{X}\,\mathbb{E}\mathbf{X}^\top \\ & = \begin{pmatrix} \mathbb{V}X_1 & \cdots & \mathbb{C}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathbb{C}(X_n, X_1) & \cdots & \mathbb{V}X_n \end{pmatrix} \end{cases} \tag{2}$$

2.2 Moments

2.3 Discrete

3 Norms and Inequalities

3.1 Cauchy-Schwartz Inequality

$$|\langle \mathbf{x}_1, \mathbf{x}_2 \rangle|^2 \leq \langle \mathbf{x}_1, \mathbf{x}_1 \rangle \cdot \langle \mathbf{x}_2, \mathbf{x}_2 \rangle \tag{3}$$

$$\mathbb{C}(X_1, X_2) \leq \mathbb{V}X_1 \cdot \mathbb{V}X_2 \tag{4}$$

3.2 Chebyshev's Inequality

$$\begin{aligned}\mathbb{P}\{g(X) \geq r\} \leq \frac{\mathbb{E}g(X)}{r} = \int_D g(x)f(x)\,dx \geq \int_{\{g(x)\geq r\}} g(x)f(x)\,dx \\ \geq r\int_{\{g(x)\geq r\}} f(x)\,dx = r\mathbb{P}\{g(X) \geq r\}.\end{aligned} \tag{5}$$

3.3 Jensen's Inequality

$$\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X) \tag{6}$$

convex function

$$\phi(tx_1 + (1-t)x_2) \leq t\phi(x_1) + (1-t)\phi(x_2) \tag{7}$$

$$ax + b \leq \phi(x) \tag{8}$$

$$\begin{aligned}\mathbb{E}\phi(X) = \int_I \phi(x)p(x)\,dx \geq \int_I (ax+b)p(x)\,dx = \\ a\int_I xp(x)\,dx + b\int_I p(x)\,dx = ax_0 + b = \phi(x_0) = \phi(\mathbb{E}X).\end{aligned} \tag{9}$$

4 Operators

4.1 Exponentiated Operators

4.1.1 Moment-Generating Functions

$$M_X(t) \equiv \mathbb{E}e^{tX} \tag{10}$$

$$M_X(t) \equiv \sum_n = 1^\infty \frac{\mathbb{E}X^n}{n!} t^n \tag{11}$$

$$\mathbb{E}X^n = \frac{d^n}{dt^n} M_X(t)|_{t=0} = M_X^{(n)}(0) \tag{12}$$

X and Y independent

$$M_{X+Y}(t) = \mathbb{E}e^{t(X+Y)} = \mathbb{E}e^{tX}e^{tY} = \mathbb{E}e^{tX}\mathbb{E}e^{tY} = M_X(t)M_Y(t) \tag{13}$$

$$M_{cX}(t) = \mathbb{E}e^{ctX} = M_X(ct) \tag{14}$$

4.1.2 Cumulants

$$K_X(t) = \ln M_X(t) \tag{15}$$

Since $\ln(1+x) = t - \frac{t^2}{2} + \cdots$

$$K_X(t) = \left(t\mathbb{E}X + \frac{t^2}{2}\mathbb{E}X^2 + \cdots\right) + \frac{1}{2}\left(t\mathbb{E}X + \cdots\right)^2 + \cdots \tag{16}$$

$$= t\mathbb{E}X + \frac{t^2}{2}\left((\mathbb{E}X)^2 - \mathbb{E}X^2\right) + \cdots \tag{17}$$

For X and Y independent

$$K_{X+Y}(t) = K_X(t) + K_Y(t) \tag{18}$$

$$K_{cX}(t) = K_X(ct) \tag{19}$$

4.1.3 Characteristic Functions

Fourier transform of probability density function

$$\phi_X(t) \equiv \mathbb{E}e^{itX} \tag{20}$$

for X and Y independent

$$\phi_{X+Y}(t) = \mathbb{E}e^{it(X+Y)} = \mathbb{E}e^{itX}e^{itY} = \mathbb{E}e^{itX}\mathbb{E}e^{itY} = \phi_X(t)\phi_Y(t) \tag{21}$$

$$\phi_{cX}(t) = \mathbb{E}e^{ictX} = \phi_X(ct) \tag{22}$$

let $Z = X + Y$ be the sum of two independent random variables

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \Rightarrow p_{X+Y}(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)\,dx \tag{23}$$

4.1.4 Extensions to Random Vectors

$$\left.\begin{matrix} X \to \mathbf{X} \\ t \to \mathbf{t} \end{matrix}\right\} \Rightarrow \begin{cases} M_{\mathbf{X}}(\mathbf{t}) \equiv \mathbb{E}e^{\mathbf{t}^\top \mathbf{X}} \\ K_{\mathbf{X}}(\mathbf{t}) \equiv \ln M_{\mathbf{X}}(\mathbf{t}) \\ \phi_{\mathbf{X}}(\mathbf{t}) \equiv \mathbb{E}e^{i\mathbf{t}^\top \mathbf{X}} \end{cases} \tag{24}$$

4.2 Transformations

4.2.1 General Transformation

$$Y = g(X) \tag{25}$$

$$\text{increasing function, } g : F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq Y = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \tag{26}$$

$$\text{decreasing function, } g : F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq Y = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)) \tag{27}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \tag{28}$$

4.2.2 Scale-location Adjustment

$$X \sim f(x) \Rightarrow \alpha + \beta X \sim \frac{1}{\beta} f(\alpha + \beta x) \tag{29}$$

5 Joint Distributions and Independence

bivariate distribution is **independent** if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \tag{30}$$

a multivariate distribution is **independent** if

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) \tag{31}$$

a multivariate distribution can be factored into **marginal** and **conditional** distributions

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)\, f_Y(y) \Rightarrow f_{X|Y} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{32}$$

if the multivariate distribution is **independent identically distributed (IID)**

$$X_1, \ldots, X_n \sim X \Rightarrow f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_X(x_i) \tag{33}$$

6 Common Functions

- Error function and complementary error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \tag{34}$$
$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) \tag{35}$$

- Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{36}$$

$$\begin{aligned}
\Gamma(x) &= \int_0^\infty t^{x-1} e^{-t}\, dt \\
&= -t^{x-1} e^{-t}\Big|_0^\infty - \int_0^\infty (x-1) t^{x-2} (-e^{-t})\, dt \\
&= (x-1) \int_0^\infty t^{x-2} e^{-t}\, dt \\
&= (x-1)\Gamma(x-1)
\end{aligned} \tag{37}$$

$$\Gamma(n) = (n-1)! \tag{38}$$

- Beta function

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt \tag{39}$$

$$\begin{aligned}
\Gamma(x)\Gamma(y) &= \int_0^\infty s^{x-1} e^{-s}\, ds \int_0^\infty t^{y-1} e^{-t}\, dt \\
&= \int_0^\infty \int_0^\infty s^{x-1} t^{y-1} e^{-(s+t)}\, ds\, dt && \begin{cases} s = uv \\ t = u(1-v) \end{cases} \\
&= \int_{u=0}^\infty \int_{v=0}^1 (uv)^{x-1} (u(1-v))^{y-1} e^{-u} u\, du\, dv && |J| = u \\
&= \int_0^\infty e^{-u} u^{x+y-1}\, du \int_0^1 v^{x-1} (1-v)^{y-1}\, dv \\
&= \Gamma(x+y)\, \mathrm{B}(x,y)
\end{aligned} \tag{40}$$

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{41}$$

$$
\begin{aligned}
\mathrm{B}\left(\frac{1}{2},\frac{1}{2}\right) &= \int_0^1 t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}}\,dt \\
&= 2\int_0^{\frac{\pi}{2}} \frac{\cos\theta\sin\theta}{\cos\theta\sin\theta}\,d\theta \\
&= \pi
\end{aligned} \tag{42}
$$

$$
\mathrm{B}\left(\frac{1}{2},\frac{1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)} \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\mathrm{B}\left(\frac{1}{2},\frac{1}{2}\right)} = \sqrt{\pi} \tag{43}
$$

- Multivariate Beta Function

$$
\mathrm{B}(\alpha_1,\ldots,\alpha_n) = \frac{\prod_{i=1}^{n}\Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{n}\alpha_i\right)} \tag{44}
$$

$$
\mathrm{B}(\alpha_1,\ldots,\alpha_n) = \frac{\Gamma(\alpha_j)\prod_{i\neq j}\Gamma(\alpha_i)}{\Gamma\left(\alpha_j+\sum_{i\neq j}\alpha_i\right)} = \frac{\prod_{i\neq j}\Gamma(\alpha_i)}{\Gamma(\sum_{i\neq j}\alpha_i)}\,\mathrm{B}\left(\alpha_j,\sum_{i\neq j}\alpha_i\right) \tag{45}
$$

7 Common Distributions

7.1 Discrete Distributions

7.1.1 Sampling With Replacement

7.1.1.1 Bernoulli

$$
\mathrm{Ber}(p) \equiv f(k|p) = p^k(1-p)^{1-k}, k \in \{0,1\} \tag{46}
$$

7.1.1.2 Binomial

$$
X_i \sim \mathrm{Ber}(p) \Rightarrow Y = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n,p) \tag{47}
$$

$$
\mathrm{Bin}(n,p) \equiv f(k|n,p) = \binom{n}{k} p^k(1-p)^{n-k}, k \in \{0,\cdots,n\} \tag{48}
$$

7.1.1.3 Negative Binomial

$$
\mathrm{NB}(k|r,p) = \mathrm{Bin}(k|k+r-1,p)\,\mathrm{Ber}(0|p) \tag{49}
$$

$$
\mathrm{NB}(k|r,p) \equiv f(k|r,p) = \binom{k+r-1}{k} p^k(1-p)^{r-k}, k \in \mathbb{N} \tag{50}
$$

7.1.1.4 Geometric

$$
\mathrm{Geo}(k|p) = \mathrm{NB}(k|1,1-p) = p(1-p)^{k-1}, k \in \mathbb{N} \tag{51}
$$

7.1.1.5 Poisson

$$\mathrm{Poi}(\lambda) = f(k|\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, k \in \mathbb{N} \tag{52}$$

$$\begin{aligned}
\mathrm{Bin}\left(n, \frac{\lambda}{n}\right) &= f\left(k \middle| n, \frac{\lambda}{n}\right) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{k}\right)^{n-k} \\
&= \frac{n \cdot n-1 \cdot \cdots \cdot n-k+1}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^k \\
&= \left[\left(\frac{n}{n}\right)\cdot\left(\frac{n-1}{n}\right)\cdot\cdots\cdot\left(\frac{n-k+1}{n}\right)\right]\cdot\left[\left(1-\frac{\lambda}{n}\right)^{-k}\right]\cdot\left[\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\right]
\end{aligned} \tag{53}$$

$$\lim_{n\to\infty}\mathrm{Bin}\left(n, \frac{\lambda}{n}\right) = \frac{\lambda^k}{k!}e^{-\lambda} \equiv \mathrm{Poi}(\lambda) \tag{54}$$

7.1.1.6 Multinomial

The multinomial distribution is realized from the sum of n repeated *dependent* Bernoulli trials, each parametrized by potentially different probabilities of individual success, p_i, and linked by the requirement that one, and only one, may be successful on any given trial, $\sum_{i=1}^k p_i = 1$:

$$\mathrm{Mul}(n, p_1, \cdots, p_k) \equiv f(x_1, \cdots, x_k|n, p_1, \cdots, p_k) = \frac{n}{\prod_{i=1}^k x_i!}\prod_{i=1}^k p_i^{x_i} = \frac{\Gamma(1+\sum_{i=1}^k x_i)}{\prod_{i=1}^k \Gamma(1+x_i)}\prod_{i=1}^k p_i^{x_i} \tag{55}$$

$$X_1, \cdots, X_k \sim \mathrm{Mul}(n, p_1, \cdots, p_k) \Rightarrow \begin{cases} \mathbb{E}X_i = np_i \\ \mathbb{V}X_i = np_i(1-p_i) \\ \mathbb{C}(X_i, X_j) = -np_ip_j, i \neq j \end{cases} \tag{56}$$

7.1.2 Sampling Without Replacement

7.1.2.1 Hypergeometric

$$\mathrm{Hyp}(n, N, K) \equiv f(k|n, N, K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{57}$$

7.1.2.2 Multivariate Hypergeometric

$$\left.\begin{aligned} \mathbf{k} &= (k_1, \ldots, k_n)^\top \\ \mathbf{K} &= (K_1, \ldots, K_n)^\top \end{aligned}\right\} \Rightarrow \mathrm{MHG}(\mathbf{k}|\mathbf{K}) = \frac{\binom{K_1}{k_1}\cdots\binom{K_n}{k_n}}{\binom{\sum_{i=1}^n K_i}{\sum_{i=1}^n k_i}} \tag{58}$$

7.2 Continuous Distributions

$$f(x|\theta) = h(x)g(\theta)e^{\eta(\theta)T(x)} \tag{59}$$

7.2.1 Gaussian

7.2.1.1 Univariate Gaussian

$$\mathrm{N}(\mu, \sigma^2) \equiv f_{\mathrm{N}}(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{60}$$

$$\mathbb{P}_{\mathrm{N}(\mu,\sigma^2)}[-\infty, x] = F_{\mathrm{N}}(x|\mu, \sigma^2) = \int_{-\infty}^{x} f_{\mathrm{N}}(x|\mu, \sigma^2)\, dx = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right) \tag{61}$$

$$\mathbb{E}(X-\mu)^{2n-1} = 0 \tag{62}$$

$$g(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\alpha\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{\alpha}} \tag{63}$$

$$\mathbb{E}(X-\mu)^{2n} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = (-2\sigma^2)^n \left.\frac{d^n}{d\alpha^n} g(\alpha)\right|_{\alpha=1} = (2n-1)!!\,\sigma^{2n} \tag{64}$$

$$M_{\mathrm{N}(\mu,\sigma^2)}(t) \equiv \mathbb{E}e^{tX} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2 - tx}{2\sigma^2}}\, dx \tag{65}$$

$$= e^{t\mu + \frac{1}{2}t^2\sigma^2} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}}\, dx \tag{66}$$

$$= e^{t\mu + \frac{1}{2}t^2\sigma^2} \tag{67}$$

$$\phi_{\mathrm{N}(\mu,\sigma^2)}(t) \equiv \mathbb{E}e^{itX} = e^{it\mu - \frac{1}{2}t^2\sigma^2} \tag{68}$$

$$\left.\begin{array}{l} X \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow M_X = e^{t\mu + \frac{1}{2}t^2\sigma^2} \\ Y \sim \mathrm{N}(\nu, \tau^2) \Rightarrow M_Y = e^{t\nu + \frac{1}{2}t^2\tau^2} \end{array}\right\} \Rightarrow M_{X+Y} = e^{t(\mu+\nu) + \frac{1}{2}t^2(\sigma^2+\tau^2)} \Rightarrow X + Y \sim \mathrm{N}(\mu+\nu, \sigma^2+\tau^2) \tag{69}$$

7.2.1.2 Standard Normal

$$Z \sim \mathrm{N}(0, 1) \tag{70}$$

7.2.1.3 Multivariate Gaussian

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \Rightarrow \begin{cases} \boldsymbol{\mu} = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{pmatrix} \\ \boldsymbol{\Sigma} = \begin{pmatrix} \mathbb{V}X_1 & \cdots & \mathbb{C}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathbb{C}(X_1, X_n) & \cdots & \mathbb{V}X_n \end{pmatrix} \end{cases} \tag{71}$$

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{72}$$

$$\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}} \tag{73}$$

Transformation, $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{Z}$,

$$\left.\begin{array}{l} \mathbb{E}\mathbf{X} \equiv \mathbb{E}(\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{Z}) = \boldsymbol{\mu} \\ \mathbb{V}\mathbf{X} \equiv \mathbb{E}\left(\mathbf{X} - \mathbb{E}\mathbf{X}\right)\left(\mathbf{X} - \mathbb{E}\mathbf{X}\right)^\top = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \end{array}\right\} \Rightarrow \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top) \tag{74}$$

$$\boldsymbol{\Sigma} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \Rightarrow \boldsymbol{\Gamma} = \mathbf{Q}\mathbf{D}^{\frac{1}{2}}, \qquad \mathbf{D}^{\frac{1}{2}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \tag{75}$$

Transformation, $\mathbf{Z} = \boldsymbol{\Gamma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$

$$\left.\begin{array}{l} \mathbb{E}\mathbf{Z} \equiv \mathbb{E}\boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{0} \\ \mathbb{V}\mathbf{Z} \equiv \mathbb{E}\left(\mathbf{Z} - \mathbb{E}\mathbf{Z}\right)\left(\mathbf{Z} - \mathbb{E}\mathbf{Z}\right)^\top = \mathbb{E}\mathbf{Z}\mathbf{Z}^\top = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-\top} = \mathbf{I} \end{array}\right\} \Rightarrow \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}) \tag{76}$$

$$\boldsymbol{\Sigma} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top \Rightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^\top \tag{77}$$

$$\mathbf{D} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2) \Rightarrow \mathbf{D}^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_n^2}\right) \tag{78}$$

$$\mathbf{y} = \mathbf{Q}^\top(\mathbf{x} - \boldsymbol{\mu}) \Rightarrow \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{2}\mathbf{y}^\top \mathbf{D}^{-1}\mathbf{y} = \sum_{i=1}^{n} \frac{y_i^2}{2\sigma_i^2} \tag{79}$$

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{y_i^2}{2\sigma_i^2}} = \prod_{i=1}^{n} f_{\mathrm{N}}(y_i|0, \sigma_i^2) \tag{80}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \Rightarrow \begin{cases} \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \\ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \end{cases} \tag{81}$$

$$\mathbf{X}_1|(\mathbf{X}_2 = \mathbf{x}_2) \sim \mathrm{N}(\boldsymbol{x}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \tag{82}$$

7.2.1.4 Marginal and Conditional Gaussian Distributions $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \Rightarrow \begin{cases} \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \\ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \end{cases} \tag{83}$$

$$(\mathbf{x}^\top - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}\,(\mathbf{x}^\top - \boldsymbol{\mu}) = \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu} \\ \mathbf{x}_2 - \boldsymbol{\mu} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu} \\ \mathbf{x}_2 - \boldsymbol{\mu} \end{pmatrix} \tag{84}$$

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I_p & 0 \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & I_q \end{pmatrix} \begin{pmatrix} \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix} \begin{pmatrix} I_p & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ 0 & I_q \end{pmatrix} \tag{85}$$

$$
\begin{aligned}
&(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\
&= \begin{pmatrix} \mathbf{x}_1-\boldsymbol{\mu} \\ \mathbf{x}_2-\boldsymbol{\mu} \end{pmatrix}^\top \begin{pmatrix} I_p \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{pmatrix} \left(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} \begin{pmatrix} I_p \\ -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}^\top \begin{pmatrix} \mathbf{x}_1-\boldsymbol{\mu} \\ \mathbf{x}_2-\boldsymbol{\mu} \end{pmatrix} + (\mathbf{x}_2-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)
\end{aligned} \tag{86}
$$

$$
\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{N}(\boldsymbol{\mu}_1-(\mathbf{x}_2-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})\ \mathrm{N}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}) \tag{87}
$$

$$
\text{Marginal Distribution: } \mathbf{X}_2 \sim \mathrm{N}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}) \tag{88}
$$

$$
\text{Conditional Distribution: } \mathbf{X}_1|(\mathbf{X}_2=\mathbf{x}_2) \sim \mathrm{N}(\boldsymbol{\mu}_1-(\mathbf{x}_2-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \tag{89}
$$

7.2.1.5 Mean and Variance of IID Normal Random Variables

$$
\left.\begin{aligned} &\mathbf{X} = (X_1,\cdots,X_n)^\top, X_i \sim \mathrm{N}(\mu,\sigma^2) \\ &\mathbf{1}_n = (1,\cdots,1)^\top \end{aligned}\right\} \Rightarrow \mathbf{X} \sim \mathrm{N}(\mu\mathbf{1}_n, \sigma^2 I_n) \tag{90}
$$

let the random vector, $\mathbf{Y}$, be formed by the linear transformaition of $\mathbf{X}$ by the $n\times n$ orthogonal matrix, Q

$$
\mathbf{Y} = (\mathbf{Y}_1,\cdots,\mathbf{Y}_n)^\top = Q^\top\mathbf{X} \Rightarrow \mathbf{Y} \sim \mathrm{N}(\mu\mathbf{1}_n, \sigma^2 Q^\top I_n Q) = \mathrm{N}(\mu\mathbf{1}_n, \sigma^2 I_n) \tag{91}
$$

Fisher's Theorem: sample mean and sample variance taken from IID normal distribution are independent:

$$
\hat{\mu}\mathbf{1}_n = \mathbf{1}_n\frac{1}{n}\mathbf{1}_n^\top\mathbf{x} = \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\mathbf{x} = P_{\mathbf{1}_n}\mathbf{x} \tag{92}
$$

$$
\hat{\sigma}^2 = \frac{1}{n-1}(\mathbf{x}-\hat{\mu}\mathbf{1}_n)^\top(\mathbf{x}-\hat{\mu}\mathbf{1}_n) = \frac{1}{n-1}(\mathbf{x}-P_{\mathbf{1}_n}\mathbf{x})^\top(\mathbf{x}-P_{\mathbf{1}_n}\mathbf{x}) = \frac{1}{n-1}\mathbf{x}^\top(I_n-P_{\mathbf{1}_n})\mathbf{x} \tag{93}
$$

These equations imply that, for arbitrary sample sets of data, information on the mean and variances of the distributed data are carried in mutually orthogonal 1- and $(n-1)$-dimensional subspaces, respectively. Knowledge of the mean carries no information about the variance, and *v.v.*

Sums of Gaussian are Gaussian

$$
\sqrt{n}\hat{\mu} \sim \mathrm{N}(\mu,\sigma^2) \tag{94}
$$

Furthermore, the distribution of the sample variance can be shown to be chi-square distributed:

$$
(n-1)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{x}^\top(I_n-P_{\mathbf{1}_n})\mathbf{x} = \sum_{i=2}^n\frac{y_i^2}{\sigma^2} = \sum_{i=2}^n z_i^2 \sim \chi^2_{n-1} \tag{95}
$$

7.2.2 Gamma-Derived Distributions

7.2.2.1 Gamma

$$
\Gamma(\alpha,\beta) \equiv f_\Gamma(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x},\ x\geq 0 \tag{96}
$$

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty t^{\alpha-1}e^{-t}\,dt \\
&= \int_0^\infty (\beta x)^{\alpha-1}e^{-\beta x}\beta\,dx \qquad t = \beta x \\
&= \int_0^\infty \beta^\alpha x^{\alpha-1}e^{-\beta x}\,dx
\end{aligned} \tag{97}$$

$$\begin{aligned}
\mathbb{E}X^n &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^k x^{\alpha-1}e^{-\beta x}\,dx \\
&= \frac{\beta^\alpha}{\beta^{\alpha+k}}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\int_0^\infty \frac{\beta^{\alpha+k}}{\Gamma(\alpha+k)}x^{\alpha+k-1}e^{-\beta x}\,dx \\
&= \frac{1}{\beta^k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\int_0^\infty f_\Gamma(x|\alpha+k,\beta)\,dx \\
&= \frac{1}{\beta^k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}
\end{aligned} \tag{98}$$

$$\mathbb{E}X = \frac{\alpha}{\beta} \tag{99}$$

$$\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{(\alpha+1)\alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2} \tag{100}$$

$$\begin{aligned}
M_{\Gamma(\alpha,\beta)} \equiv \mathbb{E}e^{tX} &= \int_0^\infty e^{tx}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\beta-t)x}\,dx \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha}\int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\beta-t)x}\,dx \\
&= \left(\frac{\beta}{\beta-t}\right)^\alpha \int_0^\infty f_\Gamma(x|\alpha,\beta-t)\,dx \\
&= \left(\frac{\beta}{\beta-t}\right)^\alpha
\end{aligned} \tag{101}$$

$$X_i \sim \Gamma(\alpha_i,\beta) \Rightarrow \sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n \alpha_i, \beta\right) \tag{102}$$

7.2.2.2 Chi-square

$$X \sim \mathrm{N}(0,1) \Rightarrow X^2 \sim \chi^2 = \Gamma\left(\frac{1}{2},\frac{1}{2}\right) \tag{103}$$

$$\mathbb{P}\{X^2 \le x\} = \mathbb{P}\{-\sqrt{x} \le X \le \sqrt{x}\} = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt \tag{104}$$

$$
\begin{aligned}
\chi^2 \equiv f_{\chi^2}(x) &= \frac{d}{dx}\int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt \\
&= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}}\right)\left(\frac{1}{2}x^{-\frac{1}{2}}\right) - \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}}\right)\left(-\frac{1}{2}x^{-\frac{1}{2}}\right) \\
&= \frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}-1} e^{-\frac{x}{2}} \\
&= \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \qquad (105)
\end{aligned}
$$

$$
X_i \sim \mathrm{N}(0,1) \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \qquad (106)
$$

$$
X \sim \chi_n^2 \Rightarrow \begin{cases} \mathbb{E}X = n \\ \mathbb{V}X = \frac{n}{2} \end{cases} \qquad (107)
$$

$$
\mathbf{\Sigma} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top = \mathbf{\Gamma}\mathbf{\Gamma}^\top \Rightarrow \mathbf{Z} = \mathbf{\Gamma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \qquad (108)
$$

$$
\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma}) \Rightarrow (\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\top \mathbf{Z} \sim \chi_n^2 \qquad (109)
$$

7.2.2.3 Inverse Gamma

7.2.3 Beta-Derived Distributions

7.2.3.1 Beta

$$
\mathrm{B}(\alpha,\beta) \equiv f_{\mathrm{B}}(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \leq x \leq 1 \qquad (110)
$$

$$
\begin{aligned}
\mathbb{E}X^k &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^k x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+\beta+k)} \int_0^1 \frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+k)\Gamma(\beta)} x^{\alpha+k-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+k)} \int_0^1 f_{\mathrm{B}}(x|\alpha+k,\beta)\,dx \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+k)} \qquad (111)
\end{aligned}
$$

$$
\mathbb{E}X = \frac{\alpha}{\alpha+\beta} \qquad (112)
$$

$$
\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2} \qquad (113)
$$

7.2.3.2 Dirichlet

$$\text{Dir}(\alpha_1, \cdots, \alpha_n) \equiv f_{\text{D}}(x_1, \cdots, x_n | \alpha_1, \cdots, \alpha_n) = \frac{\prod_{i=1}^n x_i^{\alpha_i - 1}}{\text{B}(\alpha_1, \cdots, \alpha_n)}, \begin{cases} 0 \leq x_i \leq 1 \\ \sum_{i=1}^n x_i = 1 \end{cases} \tag{114}$$

marginal distributions

$$f_D(x_i | \alpha_1, \cdots, \alpha_n) = f_{\text{B}}\left(x_i \middle| \alpha_i, \sum_{j \neq i} \alpha_j \right) = \frac{x_i^{\alpha_i - 1}(1 - x_i)^{\sum_{j \neq i} \alpha_j - 1}}{\text{B}(\alpha_i, \sum_{j \neq i} \alpha_j)} \tag{115}$$

$$\mathbb{E}X_i = \frac{\alpha_i}{\sum_{j=1}^n \alpha_j} \tag{116}$$

$$\mathbb{V}X_i = \frac{\alpha_i \sum_{j \neq i} \alpha_j}{\left(\sum_{j=1}^n \alpha_j \right)^2} \tag{117}$$

7.2.4 Distributions of Ratios of Standard Normal Random Variables

7.2.4.1 F-Distribution

$$\left. \begin{matrix} U_1, \ldots, U_k \\ V_1, \ldots, V_m \end{matrix} \right\} \sim Z = \text{N}(0, 1) \Rightarrow \frac{\frac{1}{k} \sum_{i=1}^k U_i^2}{\frac{1}{m} \sum_{j=1}^m V_j^2} \sim F(k, m),$$

$$F(k, m) \equiv f_F(x | k, m) = \frac{\Gamma(\frac{k+m}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})} \left(\frac{k}{m} \right)^{\frac{k}{2}} x^{\frac{k}{2} - 1} \left(1 + \frac{k}{m} x \right)^{-\frac{k+m}{2}} \tag{118}$$

Define the numerator and denominator in terms of chi-square distributed variables

$$\left. \begin{matrix} U = \sum_{i=1}^k U_i^2 \sim \chi_k^2 \\ V = \sum_{j=1}^m V_j^2 \sim \chi_m^2 \end{matrix} \right\} \Rightarrow \mathbb{P}\left\{ \frac{\frac{U}{k}}{\frac{V}{m}} \leq x \right\} = \mathbb{P}\{U \leq \frac{k}{m} xV\}$$

$$= \iint_{\{U \leq \frac{k}{m} xV\}} f_{\chi_k^2}(u) f_{\chi_m^2}(v) \, du \, dv = \int_0^\infty \int_0^{\frac{k}{m} xv} f_{\chi_k^2}(u) f_{\chi_m^2}(v) \, du \, dv \tag{119}$$

$$
\begin{aligned}
F(k,m) \equiv f_F(x|k,m) &= \frac{d}{dx}\int_0^\infty \int_0^{\frac{k}{m}xv} f_{\chi^2_k}(u) f_{\chi^2_m}(v)\,du\,dv = \int_0^\infty \frac{d}{dx}\left(\int_0^{\frac{k}{m}xv} f_{\chi^2_k}(u)\,du\right) f_{\chi^2_m}(v)\,dv \\
&= \frac{k}{m}\int_0^\infty f_{\chi^2_k}\left(\frac{k}{m}xv\right) f_{\chi^2_m}(v)\,v\,dv \\
&= \frac{k}{m}\int_0^\infty \left(\frac{\frac{1}{2}^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\left(\frac{k}{m}xv\right)^{\frac{k}{2}-1} e^{-\frac{\frac{k}{m}xv}{2}}\right)\left(\frac{\frac{1}{2}^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} v^{\frac{m}{2}-1} e^{-\frac{v}{2}}\right) v\,dv \\
&= \frac{k^{\frac{k}{2}}}{m}\int_0^\infty \frac{\left(\frac{1}{2}\right)^{\frac{k+m}{2}} x^{\frac{k}{2}-1}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} v^{\frac{k+m}{2}-1} e^{-\frac{1}{2}v\left(\frac{k}{m}t+1\right)}\,dv \\
&= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{k}{m}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1}\left(\frac{1}{\frac{k}{m}x+1}\right)^{\frac{k+m}{2}} \int_0^\infty \frac{\left(\frac{t+1}{2}\right)^{\frac{k+m}{2}}}{\Gamma\left(\frac{k+m}{2}\right)} v^{\frac{k+m}{2}-1} e^{-v\frac{t+1}{2}}\,dv \\
&= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{k}{m}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}} \int_0^\infty f_\Gamma\left(v\,\middle|\,\frac{k+m}{2},\frac{t+1}{2}\right) dv \\
&= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{k}{m}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}} \qquad (120)
\end{aligned}
$$

7.2.4.2 T-Distribution

$$
\left.\begin{array}{l} U_1 \\ V_1,\ldots,V_m \end{array}\right\} \sim Z = \mathrm{N}(0,1) \Rightarrow \frac{U_1}{\sqrt{\frac{1}{m}\sum_{j=1}^m V_j^2}} \sim T(m),
$$

$$
T(m) \equiv f_T(x|m) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{m}{2})}\frac{1}{\sqrt{m}}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}} \qquad (121)
$$

$$
\left.\begin{array}{l} U = U_i^2 \sim \chi^2 \\ V = \sum_{j=1}^m V_j^2 \sim \chi^2_m \end{array}\right\} \Rightarrow \mathbb{P}\left\{-x \le \frac{\sqrt{U}}{\sqrt{\frac{1}{m}V}} \le x\right\} = \mathbb{P}\left\{\frac{U}{\frac{1}{m}V} \le x^2\right\} = \int_0^{x^2} f_F(v|1,m)\,dv \qquad (122)
$$

$$
\begin{aligned}
T(m) \equiv f_T(x|m) &= \frac{1}{2}\frac{d}{dx}\int_0^{x^2} f_F(v|1,m)\,dv = x f_F\left(x^2|1,m\right) \\
&= \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{m}{2}\right)}\frac{1}{\sqrt{m}}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}} \qquad (123)
\end{aligned}
$$

7.2.4.3 Cauchy

$$
\left.\begin{array}{l} U \\ V \end{array}\right\} \sim Z = \mathrm{N}(0,1) \Rightarrow \frac{U}{V} \sim \mathrm{Cau}(0,1) \equiv f_C(x) = \frac{1}{\pi}\frac{1}{x^2+1} \qquad (124)
$$

$$
\begin{aligned}
\mathbb{P}\left\{\frac{U}{V} \le x\right\} = \mathbb{P}\left\{U \le xV\right\} &= \iint_{\{u \le xv\}} f_{\mathrm{N}}(u|0,1) f_{\mathrm{N}}(v|0,1)\,du\,dv \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{xv} f_{\mathrm{N}}(u|0,1) f_{\mathrm{N}}(v|0,1)\,du\,dv \qquad (125)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cau}(0,1) \equiv f_{\mathrm{C}}(x) &= \frac{d}{dx}\mathbb{P}\left\{\frac{U}{V} \le x\right\} = \frac{d}{dx}\int_0^\infty \int_0^{xv} f_{\mathrm{N}}(u|0,1) f_{\mathrm{N}}(v|0,1)\, du\, dv \\
&= \int_{-\infty}^\infty \left(\frac{d}{dx}\int_{-\infty}^{xv} f_{\mathrm{N}}(u|0,1)\, du\right) f_{\mathrm{N}}(v|0,1)\, dv = \int_{-\infty}^\infty f_{\mathrm{N}}(xv|0,1) f_{\mathrm{N}}(v|0,1)\, dv \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2v^2}{2}} \frac{1}{2\pi} e^{-\frac{v^2}{2}} v\, dv = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-\frac{v^2(x^2+1)}{2}} v\, dv \\
&= \frac{1}{2\pi}\frac{1}{x^2+1}\int_{-\infty}^\infty e^{-t}\, dt = \frac{1}{\pi}\frac{1}{x^2+1}\int_0^\infty e^{-t}\, dt \\
&= \frac{1}{\pi}\frac{1}{x^2+1}
\end{aligned} \tag{126}
$$

$$
\mathrm{Cau}(0,1) \equiv T(1) \tag{127}
$$

7.2.5 Other Common Distributions

7.2.5.1 Exponential

$$
\mathrm{Exp}(\lambda) \equiv f_{\mathrm{E}}(x|\lambda) = \lambda e^{-\lambda x}, \qquad x \ge 0 \tag{128}
$$

memoryless:

$$
\mathbb{P}\{x > s+t | x > s\} = \mathbb{P}\{x > t\} \tag{129}
$$

$$
h(t) = \frac{f(t)}{1 - \int_0^t f(x)\, dx} = \lambda \tag{130}
$$

7.2.5.2 Pareto

$$
\mathrm{Par}(\alpha, x_m) \equiv f_{\mathrm{P}}(x|\alpha, x_m) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, & x \ge x_m \\ 0, & x < x_m \end{cases} \tag{131}
$$

hazard rate (burn-in period)

$$
h(t) = \frac{\alpha}{t} \tag{132}
$$

7.2.5.3 Weibull

7.2.5.4 Uniform

$$
\mathrm{Uni}(0,1) \equiv f_{\mathrm{U}}(x) = 1, \;\; 0 \le x \le 1 \tag{133}
$$

8 Order Statistics

Order statistics are the probability distributions of **cumulative probability rank**, or 'percentile', of finite samples taken with replacement from arbitrary distributions.

Let a set of n IID random variables, designated as $X_1, \cdots, X_n$, be sampled from a uniform distribution, $X_i \sim \mathrm{Uni}(0,1)$. The random variables sorted in increasing order, designated as $X_{(1)}, \cdots, X_{(n)}$,

- $k-1$ events fall within $[0, u)$
- 1 event falls within $[u + du)$
- $n-k$ events fall within $[u + du, 1]$

finite-sized intervals given by multinomial probabilities,

$$\frac{n!}{(k-1)!1!(n-k)!} u^{k-1} \cdot du \cdot (1-u-du)^{n-k} \tag{134}$$

$$X_{(k)} \sim \mathrm{B}(k, n+1-k) \tag{135}$$

9 Asymptotic Limits

9.1 Convergence of Random Variables

$$\mathbb{P}\{X < x\} \equiv \mathbb{P}_X(x) \tag{136}$$

9.1.1 Convergence in Distribution

$$\lim_{n\to\infty} \mathbb{P}_{X_n}(x) = \mathbb{P}_X(x) \Leftrightarrow X_n \xrightarrow{d} X \tag{137}$$

9.1.2 Convergence in Probability (Weak Convergence)

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| \geq \epsilon\} = 0 \Leftrightarrow X_n \xrightarrow{p} X \tag{138}$$

9.1.3 Convergence Almost Surely (Strong Convergence)

$$\mathbb{P}\{\lim_{n\to\infty} |X_n - X| \leq \epsilon\} = 0 \Leftrightarrow X_n \xrightarrow{a.s.} X \tag{139}$$

9.1.4 Product of Convergent Random Variables

Slutsky's Theorem

$$\left.\begin{array}{l} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{p} c \end{array}\right\} \Rightarrow X_n Y_n \xrightarrow{d} cX \tag{140}$$

9.2 Asymptotic Limits of IID Samples

9.2.1 Law of Large Numbers

$$\mathbb{E}X = \mu \Rightarrow \mathbb{E}\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i = \mu \tag{141}$$

Weak Law of Large Numbers

$$\lim_{n\to\infty} \mathbb{P}\{|\bar{X}_n - \mu| \geq \epsilon\} = 0 \tag{142}$$

Strong Law of Large Numbers

$$\mathbb{P}\left\{\lim_{n\to\infty} |\bar{X}_n - \mu| \geq \epsilon\right\} = 0 \tag{143}$$

9.2.2 Central Limit Theorem

9.2.2.1 Univariate Theorem

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0,1) \tag{144}$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} \tag{145}$$

$$\left.\begin{array}{l} \mathbb{E}\left(\frac{X_i-\mu}{\sigma\sqrt{n}}\right) = 0 \\ \mathbb{E}\left(\frac{X_i-\mu}{\sigma\sqrt{n}}\right)^2 = \frac{1}{n} \\ \phi_X(t) = 1 + it\mathbb{E}X - \frac{t^2}{2}\mathbb{E}X^2 + \cdots \end{array}\right\} \Rightarrow \phi_{\frac{X_i-\mu}{\sigma\sqrt{n}}}(t) = 1 - \frac{t^2}{2n} + \mathcal{O}\left(n^{-\frac{3}{2}}\right) \tag{146}$$

$$\phi_{\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma\sqrt{n}}}(t) = \prod_{i=1}^{n} \phi_{\frac{X_i-\mu}{\sigma\sqrt{n}}}(t) = \prod_{i=1}^{n}\left(1 - \frac{t^2}{2n} + \mathcal{O}\left(n^{-\frac{3}{2}}\right)\right) \tag{147}$$

$$\lim_{n\to\infty}\prod_{i=1}^{n}\left(1 - \frac{t^2}{n} + \mathcal{O}\left(n^{-\frac{3}{2}}\right)\right) = \lim_{n\to\infty}\left(1 - \frac{t^2}{n} + \mathcal{O}\left(n^{-\frac{3}{2}}\right)\right)^n = e^{-\frac{t^2}{2}} = \phi_{\mathrm{N}(0,1)}(t) \tag{148}$$

9.2.2.2 Multivariate Theorem

$$\mathbb{E}\mathbf{X} = \boldsymbol{\mu} \Rightarrow \mathbb{E}\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathbf{X}_i = \boldsymbol{\mu} \tag{149}$$

$$\mathbb{V}\mathbf{X} = \mathbb{E}\mathbf{X}\mathbf{X}^\top - \mathbb{E}\mathbf{X}\,\mathbb{E}\mathbf{X}^\top = \mathbf{\Sigma} \tag{150}$$

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}) \tag{151}$$

Delta Method

Transformation $g(\mathbf{X})$

$$\left.\begin{array}{l} \mathbb{E}g(\mathbf{X}) = g(\boldsymbol{\mu}) \\ \mathbb{V}g(\mathbf{X}) = \nabla g(\boldsymbol{\mu})^\top \mathbf{\Sigma} \nabla g(\boldsymbol{\mu}) \end{array}\right\} \Rightarrow \sqrt{n}\left(g(\mathbf{X}_n) - g(\boldsymbol{\mu})\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \nabla g(\boldsymbol{\mu})^\top \mathbf{\Sigma}\nabla g(\boldsymbol{\mu})\right) \tag{152}$$

$g(\mathbf{X}) = \Gamma\mathbf{X}$

$$\left.\begin{array}{l} \mathbb{E}g(\mathbf{X}) = \Gamma\boldsymbol{\mu} \\ \mathbb{V}g(\mathbf{X}) = \Gamma^\top\mathbf{\Sigma}\Gamma \end{array}\right\} \Rightarrow \sqrt{n}\Gamma\left(\mathbf{X}_n - \boldsymbol{\mu}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \Gamma^\top\mathbf{\Sigma}\Gamma\right) \tag{153}$$

10 Likelihood Function and Information Measures

$$\mathbb{I}_S X = -\ln \mathbb{P} X \tag{154}$$

10.1 Shannon Information and Entropy

$$\mathbb{H} X = \mathbb{E} \mathbb{I}_S X \tag{155}$$

Relative Entropy or the **Kullback-Leibler Divergence**

$$D_{KL}(P||Q) = -\int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)}\, dx = \mathbb{H} X - \mathbb{E}_X \mathbb{I}_S Y \tag{156}$$

by Jensen's Inequality the relative entropy is always positive:

$$D_{KL}(P||Q) = -\int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)}\, dx \geq -\ln \left(\int_{-\infty}^{\infty} p(x) \frac{p(x)}{q(x)}\, dx \right) = -\ln \int_{-\infty}^{\infty} q(x)\, dx = 0 \tag{157}$$

Conditional Entropy

$$\mathbb{H}(X|Y) = \mathbb{H}(X, Y) - \mathbb{H} Y \tag{158}$$

Mutual Information

$$\mathbb{I}_M(X, Y) = \mathbb{H} X + \mathbb{H} Y - \mathbb{H}(X, Y) = D_{KL}\left(p(x, y) ||\, p(x) p(y)\right) = \mathbb{E}_Y D_{KL}\left(p(x|y) ||\, p(x)\right) \tag{159}$$

10.2 Likelihood Function and Fisher Information

Let $f(\mathbf{x}|\theta)$ be the joint distribution of the sample, $\mathbf{X} = (X_1, \cdots, X_n)$, taken from a distribution parametrized by θ. The **likelihood function**, $L(\theta|\mathbf{x})$, given that $\mathbf{X} = \mathbf{x}$ is observed, is defined as

$$L(\theta|\mathbf{x}) \equiv f(\mathbf{x}|\theta) \tag{160}$$

The **log-likelihood function** is identical to the Shannon information for parametrized distributions

$$\ln L(\theta|\mathbf{x}) = \ln f(\mathbf{x}|\theta) = \ln \mathbb{P} \mathbf{X}_\theta = -\mathbb{I}_S \mathbf{X}_\theta \tag{161}$$

$$L(\theta|\mathbf{x}) \equiv f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \Rightarrow \ln L(\theta|\mathbf{x}) = \sum_{i=1}^{n} \ln f(x_i|\theta) \tag{162}$$

10.2.1 Fisher Information

The **score function**, $S(\theta|\mathbf{x})$, measures the sensitivity of the likelihood function to changes in the parameter value,

$$S(\theta|\mathbf{x}) = \frac{\partial}{\partial \theta} \ln L(\theta|\mathbf{x}) = \frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta) = \frac{1}{f(\mathbf{x}|\theta)} \frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) \tag{163}$$

Two identities

$$\int_D f \frac{\partial}{\partial\theta} \ln f \, d\mathbf{x} = \int_D \frac{\partial}{\partial\theta} f \, d\mathbf{x} = \frac{\partial}{\partial\theta} \int_D f \, d\mathbf{x} = 0 \tag{164}$$

$$\int_D f \left(\frac{\partial}{\partial\theta} \ln f \right)^2 d\mathbf{x} = \int_D \frac{1}{f} \left(\frac{\partial f}{\partial\theta} \right)^2 d\mathbf{x} = \int_D \left(\frac{\partial^2 f}{\partial\theta^2} - f \frac{\partial^2}{\partial\theta^2} \ln f \right) d\mathbf{x} = -\int_D f \frac{\partial^2}{\partial\theta^2} \ln f \, d\mathbf{x} \tag{165}$$

The expection of the score function vanishes:

$$\mathbb{E}_\theta S(\theta|\mathbf{X}) = 0 \tag{166}$$

$$\mathbb{V}_\theta S(\theta|\mathbf{X}) = \mathbb{E}_\theta S(\theta|\mathbf{X})^2 - \left(\mathbb{E}_\theta S(\theta|\mathbf{X})\right)^2 = \mathbb{E}_\theta S(\theta|\mathbf{X})^2 = -\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} S(\theta|\mathbf{X}) \right) \tag{167}$$

The **Fisher information** is the variance of the score function,

$$\mathbb{I}_F \mathbf{X}_\theta = \mathbb{V} S(\theta|\mathbf{X}) = -\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} S(\theta|\mathbf{X}) \right) = -\mathbb{E}_\theta \frac{\partial^2}{\partial\theta^2} \ln f(\mathbf{X}|\theta) \tag{168}$$

Relationship between Shannon entropy and Fisher information:

$$\ln L(\theta + \Delta\theta|\mathbf{x}) \approx \ln L(\theta|\mathbf{x}) + \frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{x}) \Delta\theta + \frac{1}{2} \frac{\partial^2}{\partial\theta^2} \ln L(\theta|\mathbf{x}) \, \Delta\theta^2 \tag{169}$$

$$\begin{aligned}
-\mathbb{E}_\theta \ln L(\theta + \Delta\theta|\mathbf{X}) &\approx -\mathbb{E}_\theta \left(\ln L(\theta|\mathbf{X}) + \frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X}) \Delta\theta + \frac{1}{2} \frac{\partial^2}{\partial\theta^2} \ln L(\theta|\mathbf{X}) \Delta\theta^2 \right) \\
&= \mathbb{E}_\theta \mathbb{I}_S \mathbf{X}_\theta - \mathbb{E}_\theta S(\theta|\mathbf{X}) \Delta\theta + \frac{1}{2} \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} S(\theta|\mathbf{X}) \right) \Delta\theta^2 \\
\mathbb{H} \mathbf{X}_{\theta+\Delta\theta} &\approx \mathbb{H} \mathbf{X}_\theta + \frac{1}{2} \mathbb{I}_F \mathbf{X}_\theta \Delta\theta^2
\end{aligned} \tag{170}$$

Kullback-Leibler divergence

$$\begin{aligned}
D_{KL} \left(L(\theta|\mathbf{x}) || L(\theta + \Delta\theta|\mathbf{x}) \right) &= \int_\Theta L(\theta|\mathbf{x}) \ln \frac{L(\theta|\mathbf{x})}{L(\theta + \Delta\theta|\mathbf{x})} \, d\theta \\
&= \int_\Theta L(\theta|\mathbf{x}) \left(\ln L(\theta|\mathbf{x}) - \ln L(\theta + \Delta\theta|\mathbf{x}) \right) \, d\theta \\
&\approx \int_\Theta L(\theta|\mathbf{x}) \left(-\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{x}) \Delta\theta - \frac{1}{2} \frac{\partial^2}{\partial\theta^2} \ln L(\theta|\mathbf{x}) \Delta\theta^2 \right) \, d\theta \\
&= -\mathbb{E}_\theta S(\theta|\mathbf{X}) \, \Delta\theta - \frac{1}{2} \mathbb{E}_\theta \frac{\partial}{\partial\theta} S(\theta|\mathbf{X}) \, \Delta\theta^2 \\
&= \frac{1}{2} \mathbb{I}_F \mathbf{X}_\theta \, \Delta\theta^2
\end{aligned} \tag{171}$$

Given a vector of parameters, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^\top$, the **Fisher information matrix** is givenby

$$\mathbb{I}_F \mathbf{X}_{\boldsymbol{\theta}} \equiv -\mathbb{E}_{\boldsymbol{\theta}} \nabla^2 \ln f(\mathbf{X}|\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}} \begin{pmatrix} \frac{\partial^2}{\partial\theta_1^2} \ln f(\mathbf{X}|\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_1 \partial\theta_m} \ln f(\mathbf{X}|\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial\theta_m \partial\theta_1} \ln f(\mathbf{X}|\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_m^2} \ln f(\mathbf{X}|\boldsymbol{\theta}) \end{pmatrix} \tag{172}$$

Multidimensional Taylor expansion of the log-likelihood function:

$$\ln L(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}|\mathbf{x}) \approx \ln L(\boldsymbol{\theta}|\mathbf{x}) + \Delta\boldsymbol{\theta}^\top \ln L(\boldsymbol{\theta}|\mathbf{x}) + \frac{1}{2}\Delta\boldsymbol{\theta}^\top H \ln L(\boldsymbol{\theta}|\mathbf{x})\Delta\boldsymbol{\theta} \tag{173}$$

$$D_{KL}\left(L(\boldsymbol{\theta}|\mathbf{x})||L(\boldsymbol{\theta}+\Delta\boldsymbol{\theta}|\mathbf{x})\right) = \Delta\boldsymbol{\theta}^\top \frac{1}{2}\mathbb{I}_F \mathbf{X}_{\boldsymbol{\theta}}\,\Delta\boldsymbol{\theta} \tag{174}$$

11 Bayesian Perspectives

11.1 Bayes' Theorem

Given events, A and B, and a probability measure, $\mathbb{P}$, **Bayes' Theorem** is

$$\mathbb{P}\{A|B\} = \frac{\mathbb{P}\{B|A\}\mathbb{P}\{A\}}{\mathbb{P}\{B\}} \tag{175}$$

11.2 Parameter Refinement and Estimation

the object is to estimate the unknown value of a parameter, θ^*, that governs a parametrized distribution, $f(x|\theta^*)$. The estimate for the parameter is made through a sequence of IID measurements, $x_1, \cdots, x_i$, sampled from the distribution, and each of which adds to the refinement of a distribution for the parameter, $f(\theta)$.

11.3 Conjugate Families

Beta prior, binomial likelihood (n draws):

$$\left.\begin{array}{ll}\text{beta prior:} & \mathrm{B}(\alpha,\beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ \text{binomial likelihood:} & \mathrm{Bin}(n,k) \propto \theta^k(1-\theta)^{n-k}\end{array}\right\} \Rightarrow \text{beta posterior: } \mathrm{B}(\alpha+k, \beta+n-k) \tag{176}$$

Gamma prior, Posson likelihood (n draws):

$$\left.\begin{array}{ll}\text{gamma prior:} & \Gamma(\alpha,\beta) \propto x^{\alpha-1}e^{-\beta x} \\ \text{Poisson likelihood:} & \mathrm{Poi}(x) \propto x^k e^{-nx}\end{array}\right\} \Rightarrow \text{gamma posterior: } \Gamma(\alpha+k, \beta+n) \tag{177}$$

Gaussian prior, Gaussian likelihood (n draws): This is a model for an iterative determination of the unknown mean of a Gaussian distribution with known variance:

$$\left.\begin{array}{ll}\text{Gaussian prior:} & \mu \sim \mathrm{N}(\mu_0,\sigma_0) \propto \exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right) \\ \text{Gaussian likelihood:} & \mu|\mathbf{x} \sim \mathrm{N}(\mu,\sigma^2) \propto \prod_{i=1}^n \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)\end{array}\right\}$$
$$\Rightarrow \text{Gaussian posterior: } \mu \sim \mathrm{N}\left(\frac{\kappa_0\mu_0 + n\bar{\mathbf{x}}}{\kappa_0+n}, \frac{\sigma^2}{\kappa_0+n}\right) \tag{178}$$

$$\kappa_0 = \frac{\sigma}{\sigma_0} \tag{179}$$

Inverted gamma prior, Gaussian likelihood This is a model for an iterative determination of the unknown variance of a Gaussian distribution with known mean

$$\left.\begin{array}{rl} \text{inverted gamma prior:} & \sigma^2 \sim \mathrm{I}\,\Gamma(\alpha,\beta) \propto \sigma^{-2(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right) \\ \text{Gaussian likelihood:} & \sigma^2|\mathbf{x} \sim \mathrm{N}(\mu,\sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right) \end{array}\right\}$$
$$\Rightarrow \text{inverted gamma posterior: } \sigma^2 \sim \mathrm{I}\,\Gamma\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \tag{180}$$

11.4 Monte Carlo Methods