Section 1 Preliminaries
Solutions

Vera Liu

February 12, 2018

1 1.1

1.1

Clearly, the sequence is not iid because the number of balls of each color for future draws depends on the color of previous draws. However, the sequence is exchangeable, in the sense that if we pick up R red balls and B blue balls for any integer R and B (from an urn that originally contain r red balls and b blue balls),

$P(X_1, X_2...X_{R+B}) = \frac{\frac{(R+r)!(B+b)!}{(r-1)!(b-1)!}}{\frac{(R+B+r+b)!}{(r+b-1)!}} = \frac{(R+r)!(B+b)!(r+b-1)!}{(R+B+r+b)!(r-1)!(b-1)!}$, regardless of the order.

1.2

Since the sequence is exchangeable, each sequence of events has equal probability. Thus,

$P(X_1 = x_1, X_2 = x_2...X_N = x_N | \sum_{i=1}^{M} x_i = t) = \frac{1}{\binom{M}{t}}$.

$P(\sum_{i=1}^{N} x_i = s | \sum_{i=1}^{M} x_i = t) = \sum P(X_1 = x_1, X_2 = x_2...X_M = x_M | \sum_{i=1}^{M} x_i = t) = \frac{\binom{N}{s}\binom{M-N}{t-s}}{\binom{M}{t}}$

1.3

$P(\sum_{i=1}^{N} x_i = s) = \binom{N}{s} \int_0^1 \frac{(M\theta)_s(M(1-\theta))_{n-s}}{(M)_N} dF_M(\theta) = \binom{N}{s} \int_0^1 \frac{(M\theta)!}{(s)!} \frac{(M(1-\theta))!}{(n-s)!} \frac{N!}{M!} dF_M(\theta) \approx \binom{N}{s} \int_0^1 \frac{(M\theta)^s(M(1-\theta))^{N-s}}{M^N} dF_M(\theta)$ with $M \to \infty$

Thus, above equation becomes $\binom{N}{s} \int_0^1 \theta^s((1-\theta))^{N-s} dF_M(\theta)$

1.4

For Poisson random variable X, $p(x|\lambda) = (\frac{\lambda^x e^{-\lambda}}{x!}) = \frac{exp(xlog\lambda - \lambda)}{x!}$ is of exponential family, with $h(x) = x!$, $T(x) = x$, $\nu(\theta) = log\lambda$, $A(\nu(\theta)) = \lambda = e^{\nu}$

1.5

Assuming $X_1...X_N$ are iid gamma random variables, $P(X|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}^N (X_1 \times ... \times X_N) exp(-\beta(X_1 + ... + X_N)) = exp(\alpha N log\beta - N log(\Gamma(\alpha)) + (\alpha - 1) log(X_1 \times ... \times X_N) - \beta(X_1 + ... + X_N)$ is of exponential family, with $h(x) = 1$, $T(x) = [log((X_1 \times ... \times X_N)), (X_1 + ... + X_N)]^T$, $\nu(\alpha, \beta) = [\alpha - 1, \beta]^T$, $A(\nu) = \alpha N log\beta - N log(\Gamma(\alpha))$

1.6

For exponential family distributions with density $f(x) = h(x) exp(\nu(\theta)T(x) - A(\nu))$,

$M_{T(X)}(s) = E(e^{sT(x)}|\nu) = \int_{-\infty}^{\infty} h(x) exp(\nu(\theta)T(x) - A(\nu)) exp(sT(x)) dx = exp(A(\nu + s) - A(\nu)) \int_{-\infty}^{\infty} h(x) exp((s + \nu(\theta))T(x) - A(\nu + s)) dx = exp(A(\nu + s) - A(\nu))$

1.7

As shown in previous problems, in Poisson distribution, $A(\nu) = e^\nu$

Hence $M_{T(X)}(s) = exp(e^{s+\nu} - e^\nu)$

$E(X) = \frac{\partial exp(e^{s+\nu} - e^\nu))}{\partial s} = e^{s+\nu} exp(e^{s+\nu} - e^\nu)) = e^\nu = \lambda$ when evaluating at $s = 0$

$E(X^2) = \frac{\partial\partial exp(e^{s+\nu} - e^\nu))}{\partial\partial s} = e^{s+\nu} exp(e^{s+\nu} - e^\nu)) + e^{s+\nu} e^{s+\nu} exp(e^{s+\nu} - e^\nu)) = e^\nu + e^\nu e^\nu = \lambda^2 + \lambda$ when evaluating at $s = 0$.

$Var(X) = E(X^2) - E(X)^2 = \lambda$

$C_{T(X)}(s) = e^{s+\nu} - e^\nu$

$E(X) = \frac{\partial C_{T(X)}(s)}{\partial s} = e^{s+\nu} = e^\nu = \lambda$ when evaluating at $s = 0$

$Var(X) = \frac{\partial\partial C_{T(X)}(s)}{\partial\partial s} = e^{s+\nu} = e^\nu = \lambda$ when evaluating at $s = 0$

1.8

$L(X|\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^N e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

$f(\mu) = (2\pi{\sigma_0}^2)^{-\frac{1}{2}} e^{-\frac{(\mu-\mu_0)^2}{2{\sigma_0}^2}}$

$f(\mu|X) \propto L(X|\mu) \times f(\mu)$

$\propto e^{-\sum_{i=1}^N \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2{\sigma_0}^2}}$

$\propto e^{-\frac{2\sigma^2{\sigma_0}^2}{N{\sigma_0}^2+\sigma^2} \times (\mu - \frac{{\sigma_0}^2 \sum_{i=1}^N x_i + \sigma^2\mu_0}{N{\sigma^2}_0 + \sigma^2})^2}$

Hence $\mu|X \sim N(\frac{{\sigma_0}^2 \sum_{i=1}^N x_i + \sigma^2\mu_0}{N{\sigma^2}_0+\sigma^2}, \frac{\sigma^2{\sigma_0}^2}{n{\sigma_0}^2+\sigma^2})$

If in terms of precision, i.e. Let $\tau_0 = \frac{1}{{\sigma_0}^2}$ and $\tau = \frac{1}{\sigma^2}$, then $\mu|X \sim N(\frac{\tau_0\mu_0 + \tau\sum_{i=1}^N x_i}{\tau_0 + N\tau}, \tau_0 + N\tau)$

1.9

$L(X|\omega) \propto \omega^{\frac{1}{2}} \prod_{i=1}^N e^{-\frac{\omega}{2}(x_i-\mu)^2}$

$\propto \omega^{\frac{N}{2}} e^{-\frac{\omega}{2}\sum_{i=1}^N (x_i-\mu)^2}$

$f(\omega) = \frac{\beta^\alpha}{\Gamma(\alpha)} \omega^{\alpha-1} e^{-\beta\omega}$

$f(\omega|X) \propto L(X|\omega) \times f(\omega)$

$\propto \omega^{\alpha-1+\frac{N}{2}} \times e^{-\beta\omega - \frac{1}{2}\omega\sum_{i=1}^N (x_i-\mu)^2}$

Hence $\omega|X \sim Gamma(\alpha + \frac{N}{2}, \beta + \sum_{i=1}^N (x_i - \mu)^2)$

1.10

We have the joint distribution:

$f(x, \omega) = L(x|\omega) \times f(\omega)$

$= \frac{\beta^\alpha}{\Gamma(\alpha)} \sqrt{\frac{1}{2\pi}} \omega^{\frac{1}{2}+\alpha-1} e^{-\beta\omega - \frac{\omega}{2}x^2}$

To get marginal distribution:

$f(x) = \int_{-\infty}^{\infty} f(x, \omega) d\omega$

$= \int_{-\infty}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \sqrt{\frac{1}{2\pi}} \omega^{\frac{1}{2}+\alpha-1} e^{-\beta\omega - \frac{\omega}{2}x^2} d\omega$

$\omega^{\frac{1}{2}+\alpha-1} e^{-\beta\omega - \frac{\omega}{2}x^2}$ is the kernel of $Gamma(\frac{1}{2} + \alpha, \beta + \frac{1}{2}x^2)$, hence

$\int_{-\infty}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \sqrt{\frac{1}{2\pi}} \omega^{\frac{1}{2}+\alpha-1} e^{-\beta\omega - \frac{\omega}{2}x^2} d\omega$

$= \frac{\beta^\alpha}{\Gamma(\alpha)} \sqrt{\frac{1}{2\pi}} \frac{\Gamma(\alpha + \frac{1}{2})}{(\beta + \frac{1}{2}x^2)^{\frac{1}{2}+\alpha}}$

$= \frac{\Gamma(\frac{1}{2}+\alpha)}{\Gamma(\alpha)\sqrt{2\pi\beta}} (1 + \frac{x^2}{2\beta})^{-\frac{1}{2}-\alpha}$

is the density of a non-central $t-$ distribution

1.11

$\Sigma = E((X-\mu)(X-\mu)^T) = E(XX^T) - 2\mu E(X^T) + \mu\mu^T = E(XX^T) - \mu\mu^T$

$cov(AX+b) = E((AX-A\mu)(AX-A\mu)^T) = E(A(X-\mu)(X-\mu)^T A^T) = AE((X-\mu)(X-\mu)^T)A^T = A\Sigma A^T$

1.12

Let $X \sim N(\mu, \sigma^2)$, $M_X(t) = E(e^{tX}) = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} e^{tx} dx = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

When X is a standard multivariate normal distribution in vector form,

$M_X(t) = E(e^{t^T X}) = E(exp(\sum_i t_i X_i) = E(\prod_i exp(t_i X_i)) = \prod_i E(exp(t_i X_i))$ since X_i are independent of each other in standard multivariate normal. Then plug in the moment generating function for univariate normal as each X_i is a univariate normal random variable. $M_X(t) = e^{\frac{1}{2}t^T t}$

1.13

X is multivariate normal $\Leftrightarrow$ $a^T X$ is univariate normal $\Leftrightarrow$ $M_{a^T X}(s) = E(e^{sa^T X}) = exp(sa^T\mu + \frac{1}{2}s^T a^T \Sigma a s) = exp(s(a^T\mu) + \frac{1}{2}s^T(a^T\Sigma a)s)$, $\Leftrightarrow M_X(s) = exp(s\mu + \frac{1}{2}s^T\Sigma s)$

1.14

As shown in previous problem, X has a multivariate normal distribution $\Leftrightarrow M_X(s) = exp(s^T\mu + \frac{1}{2}s^T\Sigma s) = exp(s^T\mu + \frac{1}{2}s^T D^T D s)$ (because Σ is a positive semi-definite matrix, so there must exist such matrix D).

$\Leftrightarrow M_X(s) = M_{\mu+DZ}(s)$ because $M_Z(s)$ takes the form of $exp(\frac{1}{2}z^T z$

Any multivariate normal vectors of length N could be generated by first generating the standard multivariate normal vector of the same length, and left multiply by D, where D satisfies $D^T D = \Sigma$, and then add the mean vector.

1.15

For standard multivariate normal variable Z, $f(z) = (\frac{1}{2\pi})^{\frac{n}{2}} e^{-\frac{z^T z}{2}}$.

Using change-of-variables, let $X = DZ + \mu$, $Z = D^{-1}(X-\mu)$, Jacobian $= |D^{-1}| = |\Sigma|^{-\frac{1}{2}}$.

Substitute the new variable and Jacobian to the standard normal density function, we have

$f(x) = (\frac{1}{2\pi})^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{(x-\mu)^T D^{-1} D^{-1}(x-\mu)}{2}}$

$= (\frac{1}{2\pi})^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}}$

1.16

Since x is a multivariate normal, $a^T x$ is a univariate normal for $\forall a$. Pick $a = (1,0)^T$, then we have $x_1 = (1,0)^T x$ is a univariate normal. This works for all a where the coefficient for x_2 is 0. Hence the marginal distribution of x_1 is a multivariate normal.

$E(x_1) = E((1,0)^T x) = (1,0)^T \mu = \mu_1$

$Cov(x_1) = Cov((1,0)^T x) = (1,0)\Sigma(1,0)^T = \Sigma_{11}$

$X_1 \sim N(\mu_1, \Sigma_{11})$

1.17

$\Sigma\Omega = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix} \times \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{bmatrix}$

Hence we have equations:

$$\begin{cases} \Sigma_{11}\Omega_{11} + \Sigma_{12}\Omega_{12}^T = I \\ \Sigma_{11}\Omega_{12} + \Sigma_{12}\Omega_{22} = 0 \\ \Sigma_{12}^T\Omega_{11} + \Sigma_{22}\Omega_{12}^T = 0 \\ \Sigma_{12}^T\Omega_{12} + \Sigma_{22}\Omega_{22} = I \end{cases}$$

Solve the equations, we get:
$$\begin{cases}\Omega_{12} = (\Sigma_{12}^T + \Sigma_{22}\Sigma_{12}^{-1}\Sigma_{11})^{-1} \\ \Omega_{11} = (\Sigma_{12}^T)^{-1}\Sigma_{22}(\Sigma_{12}^T + \Sigma_{22}\Sigma_{12}^{-1}\Sigma_{11})^{-1} \\ \Omega_{22} = \Sigma_{22}^{-1} - \Sigma_{22}^{-1}\Sigma_{12}^T(\Sigma_{12}^T + \Sigma_{22}\Sigma_{12}^{-1}\Sigma_{11})^{-1}\end{cases}$$

1.18

$cov(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2, X_2) = cov(X_1, X_2) - cov(\Sigma_{12}\Sigma_{22}^{-1}X_2, X_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$

$X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, cov(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2))$then by independence of $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ and X_2, $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2|X_2 \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, cov(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2))$.

$cov(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2))$
$= \mathrm{E}[(\mathrm{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2)(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2)^T] - (\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2)(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2)^T$
$= [\mathrm{E}(\mathrm{X}_1 X_1^T) - \mu_1\mu_1^T] + [E(\Sigma_{12}\Sigma_{22}^{-1}X_2X_2^T\Sigma_{22}^{-1}\Sigma_{12}^T) - (\Sigma_{12}\Sigma_{22}^{-1}\mu_2\mu_2^T\Sigma_{22}^{-1}\Sigma_{12}^T)] - 2[E(X_1X_2^T\Sigma_{22}^{-1}\Sigma_{12}^T) - \mu_1\mu_2^T\Sigma_{22}^{-1}\Sigma_{12}^T]$
$=\Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$
$=\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$

Hence $X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)$

1.19

For $\forall d = 1, 2, ...p$, $0 = cov(X_d, \epsilon) = E((X_d - \bar{X_d})^T(\epsilon - \bar{\epsilon})) = E({X_d}^T\epsilon - {X_d}^T\bar{\epsilon} - \bar{X_d}^T\epsilon + \bar{X_d}^T\bar{\epsilon})$, plug in $\epsilon = Y - X\beta$, we have $E({X_d}^TY - {X_d}^TX\beta) = E(\bar{X_d}^TY - \bar{X_d}^TX\beta)$. $\hat{\beta_{MM}} = ({X_d}^TX)^{-1}{X_d}^TY$ and this is the coefficient estimate for each covariate. Thus we have $\hat{\beta_{MM}} = (X^TX)^{-1}X^TY$

1.20

$y_i \sim N(x_i\beta, \sigma^2)$.

Likelihood is $L(Y) = (2\pi)^{-\frac{n}{2}}exp(-\frac{(y_i - x_i\beta)^T(y_i - x_i\beta)}{2\sigma^2})$

Maximizing the likelihood is equivalent as minimizing the negative log-likelihood, i.e.minimizing $(Y - X\beta)^T(Y - X\beta)$.

Set $\frac{\partial L(Y)}{\partial\beta} = 0 = X^T(Y - X\beta)$

Hence $\hat{\beta_{ML}} = (X^TX)^{-1}X^TY$

1.21

For least square estimate, we'll also be minimizing $(Y - X\beta)^T(Y - X\beta)$, which is essentially the same as under ML estimates. Hence $\hat{\beta_{ML}} = \hat{\beta_{OLS}}$

1.22

$l(\beta) = (y - x\beta)^T(y - x\beta) + \lambda(\beta^T\beta - t)$. Taking derivative to optimize, $\partial l(\beta) = -2x^T(y - x\beta) + 2\lambda\beta = 0 \Leftrightarrow \hat{\beta}_{Ridge} = (X^TX + \lambda I)^{-1}X^TY = (XX^T + \lambda I)^{-1}X^TX\hat{\beta}_{OLS}$

1.23

$\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY = (X^TX)^{-1}X^T(X\beta + \epsilon) = \beta + (X^TX)^{-1}X^T\epsilon$

$E(\hat{\beta}_{OLS}) = E(\beta + (X^TX)^{-1}X^T\epsilon) = \beta$ because $E(\epsilon) = 0$ and is independent of X

$Var(\hat{\beta}_{OLS}) = E((\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)^T) = E((X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}) = \sigma^2(X^TX)^{-1}$

The sampling distribution of $\hat{\beta}_{OLS}$ is $N(\beta, \sigma^2(X^TX)^{-1})$

1.24

If assuming X is orthonormal, i.e.$X^T X = I$, we have $\hat{\beta}_{Ridge} = \frac{1}{1+\lambda}\hat{\beta}_{OLS}$

The sampling distribution of $\hat{\beta}_{Ridge}$ is $N((X^T X + \lambda I)^{-1} X^T X\beta, \sigma^2 (X^T X + \lambda I)^{-1}(X^T X)(X^T X + \lambda I)^{-1})$

$MSE(\hat{\beta}_{Ridge}) = \sigma^2 tr(W_\lambda (X^T X)^{-1} W_\lambda^T) + \beta(W_\lambda - I)^T (W_\lambda - I)\beta$

$MSE(\hat{\beta}_{OLS}) = tr(\sigma^2 X^T X)$

where $W_\lambda = (X^T X + \lambda I)^{-1} X^T X$

1.25

The mean square error $MSE = \frac{SSE}{df}$ is an unbiased estimator ot σ^2. Here $df = 102 - 4$, (number of observations - number of predictors).

$\hat{\sigma}^2 = \frac{\sum(Y_i - \hat{Y}_i)}{98} = 7.846$, is exactly the same as the lm output from R.

Lastly, the variance of $\hat{\beta_{OLS}}$ could be estimated by $\sigma^2 (X^T X)^{-1}$

1.26

$Var(f(\theta)) = Var(\sum_{i=1}^{p} \theta_i) = \sum_{i=1}^{p} \sum_{j=1}^{p} cov(\theta_i, \theta_j) = \sum_{i=1}^{p} Var(\theta_i) + 2 \times \sum_{1<i<j\leq p} cov(\theta_i, \theta_j)$

1.27

Using Taylor expansion on $\theta = \hat{\theta}$,

$f(\theta) = f(\hat{\theta}) + f^{'}((\hat{\theta})(\theta - \hat{\theta})O(\theta - \hat{\theta})^2$,

$Var(f(\theta)) = Var(f(\hat{\theta})) + Var(f^{'}(\hat{\theta})(\theta - \hat{\theta})) = Var(f^{'}(\hat{\theta})(\theta - \hat{\theta})) = f^{'}(\hat{\theta})Var(\theta - \hat{\theta})(f^{'}(\hat{\theta}))^T = f^{'}(\hat{\theta})\Sigma(f^{'}(\hat{\theta}))^T$