Stat 512
Homework key 4

November 2, 2015

REGULAR PROBLEMS

1. In this problem we consider alternative approaches to handle initial measurements on subjects used in a randomized experiment

   - We consider a problem in which we randomly select n independent pairs of two independent individuals from a population, so all individuals are totally independent. (In terms of statistical design, the pairs constitute "blocks".)
   - On each individual, we make initial measurements of some random variable X_{ik} having mean μ and variance σ^2 for $i = 1, 2$ and $1 \leq k \leq n$.
   - We then apply treatment A to the first chosen individual of each pair, and again measure the random variable which for notational convenience we label Y_{1k} having mean $\mu + \delta_A$ and variance σ^2. (So X and Y represent the same scientific quantity measured at different times.) We apply treatment B to the second chosen individual, and make measurement Y_{2k} having mean $\mu + \delta_B$ and variance σ^2. (In statistical design, we would randomize which individual of each block receives treatment A and which receives treatment B.)
   - We consider that repeat measurements being made on the same individual are correlated, but measurements made on different individuals are independent, with $\text{corr}(X_{ik}, Y_{i'k'}) = \rho \mathbf{1}_{[i=i' \& k=k']}$ for $i = 1, 2$.

   Ultimately, we are interested in the difference in treatment effects as measured by $\theta = \delta_A - \delta_B$.

   (a) Write down the mean and variance of random vector $W_k = (X_{1k}, Y_{1k}, X_{2k}, Y_{2k})^T$.

   (b) Define $Z_k^{(0)} = Y_{1k} - Y_{2k}$. Express $Z_k^{(0)}$ as a linear transformation of W_k for some vector $\vec{c}_{(0)}$ (so $Z_k^{(0)} = \vec{c}_{(0)} W_k$), and provide the mean and variance of $Z_k^{(0)}$. (This approach completely ignores the initial measurements.)

   (c) Define differences $D_{ik} = Y_{ik} - X_{ik}$ for $i = 1, 2$ and $1 \leq k \leq n$. Find the mean and variance for D_{ik}. Now define $Z_k^{(1)} = D_{1k} - D_{2k}$. Express $Z_k^{(1)}$ as a linear transformation of W_k for some vector $\vec{c}_{(1)}$ (so $Z_k^{(1)} = \vec{c}_{(1)} W_k$), and provide the mean and variance of $Z_k^{(1)}$. (This approach considers changes in the measurements.)

   (d) For some specified a, define the transformations $G_{ik} = Y_{ik} - aX_{ik}$ for $i = 1, 2$ and $1 \leq k \leq n$. Find the mean and variance for G_{ik}. Now define $Z_k^{(a)} = G_{1k} - G_{2k}$. Express $Z_k^{(a)}$ as a linear transformation of W_k for some vector $\vec{c}_{(a)}$ (so $Z_k^{(a)} = \vec{c}_{(a)} W_k$), and provide the mean and variance of $Z_k^{(a)}$. (This approach completely considers the arbitrary linear handling of the initial measurements.) In what sense is this result a generalization of the previous two approaches.

   (e) Now consider how averages across blocks might serve as estimators of θ. That is, consider the distributions of

   $$\bar{Z}^{(*)} = \frac{1}{n} \sum_{k=1}^{n} Z_k^{(*)}$$

   in terms of their means and variances. Find the value of a such that $\bar{Z}^{(a)}$ would have the greatest precision in terms of variance. Find the value of a such that $\bar{Z}^{(a)}$ would have the lowest mean squared error (MSE) as an estimator of θ, where

   $$MSE_\theta = E\left[(\bar{Z}^{(a)} - \theta)^2\right].$$

Ans:

(a) The mean vector of W_k follows component-wise,

$$EW_k = E(X_{1k}, Y_{1k}, X_{2k}, Y_{2k})^T = (\mu, \mu + \delta_A, \mu, \mu + \delta_B)^T.$$

The variance-covariance matrix of W_k is the block diagonal matrix given below. This structure follows directly from the given correlation and independence described in the problem.

$$V_W = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \rho & 1 \end{pmatrix}.$$

(d) We proceed by directly solving part (d) and showing that (b) and (c) follow as special cases of the result. Recalling that $G_{ik} = Y_{ik} - aX_{ik}$, we have that

$$EG_{1k} = \mu + \delta_A - a\mu = \delta_A + \mu(1-a), \quad EG_{2k} = \delta_B + \mu(1-a).$$

Then, the variances are given by

$$V_G \equiv Var(G_{1k}) = Var(G_{2k}) = \sigma^2 + a^2\sigma^2 - 2a\rho\sigma^2 = \sigma^2(1 + a^2 - 2a\rho).$$

Clearly, for $a = 0$, $G_{ik} \equiv Y_{ik}$ as used in (a). For $a = 1$, instead $G_{ik} = D_{ik}$ as defined in (b).

Now we can write $Z_k^{(a)}$ as the linear transformation

$$Z_k^{(a)} = G_{1k} - G_{2k} = 1 \cdot Y_{1k} + (-a) \cdot X_{1k} + (-1) \cdot Y_{2k} + a \cdot X_{2k} = (-a, 1, a, -1)W_k \equiv \vec{c}_{(a)}^{\,T} W_k.$$

Here, $\vec{c}_{(a)} = (-a, 1, a, -1)^T$ is defining our linear transformation. This allows us to use the linearity properties of expectation and quadratic properties of variances as follows:

$$EZ_k^{(a)} = E\vec{c}_{(a)}^{\,T} W_k = \vec{c}_{(a)}^{\,T} EW_k = (-a, 1, a, -1)(\mu, \mu + \delta_A, \mu, \mu + \delta_B)^T = \delta_A - \delta_B.$$

$$\sigma_Z^2 \equiv \sigma^2_{Z_k^{(a)}} = \vec{c}_{(a)}^{\,T} V_W \vec{c}_{(a)} = 2V_G = 2\sigma^2(1 + a^2 - 2a\rho).$$

We summarize the cases for parts (b) and (c) below.

a	G_{ik}	EG_{ik}	$Var(G_{ik})$	$\vec{c}_{(a)}^{\,T}$	$EZ_k^{(a)}$	$Var(Z_k^{(a)})$
0	Y_{ik}	$\delta_i + \mu$	σ^2	$(0, 1, 0, -1)$	$\delta_A - \delta_B$	$2\sigma^2$
1	D_{ik}	δ_i	$2\sigma^2(1-\rho)$	$(-1, 1, 1, -1)$	$\delta_A - \delta_B$	$4\sigma^2(1-\rho)$

(e) We write $\bar{Z}^{(a)} = \frac{1}{n}\sum_{k=1}^n Z_k^{(a)}$. Since the blocks are assumed independent (by our totally independent sampling of individuals), we have that

$$Var(\bar{Z}^{(a)}) = Var(\frac{1}{n}\sum_{k=1}^n Z_k^{(a)}) = \frac{1}{n^2}\sum_{k=1}^n \sigma_Z^2 = \frac{1}{n}2\sigma^2(1 + a^2 - 2a\rho).$$

Further, since $E\bar{Z}^{(a)} = \frac{1}{n} nEZ_k^{(a)} = \delta_A - \delta_B = \theta$, we have that $MSE_\theta = Var(\bar{Z}^{(a)})$. Hence, the most precise estimator in the family $\bar{Z}^{(a)}$ will be determined by a minimizing $\frac{1}{n}2\sigma^2(1 + a^2 - 2a\rho)$. This is equivalent to just minimizing

$$a^2 - 2a\rho.$$

The derivative of this function is

$$\frac{d}{da}(a^2 - 2a\rho) = 2a - 2\rho$$

which has a root at $a = \rho$. Since $a^2 - 2a\rho$ is convex:

$$\frac{d^2}{da^2}(a^2 - 2a\rho) = 2 > 0,$$

this is indeed a (in fact, *the*) minimizing value. For $a = \rho$, the variance of the estimator $\bar{Z}^{(a)}$ is just

$$Var(\bar{Z}^{(a)}) = \frac{1}{n}2\sigma^2(1 - \rho^2).$$

We will see similar arguments applied when we derive more general *best linear unbiased estimators (BLUE)*.

2. For each of the following distributions derive the mean, variance, skewness, and kurtosis.

(a) Poisson: $X \sim \mathcal{P}(\lambda)$

(b) Exponential: $X \sim \mathcal{E}(\lambda)$

(c) Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$

Ans: We illustrate below several useful patterns that simplify the calculation of moments somewhat in each case. It is not expected that these exact results were used in student solutions. The purpose of the solution below is to show alternatives to direct calculation of $E(X^k)$ or $\mu_k = E(X-\mu)^k$ for $k = 1, 2, 3, 4$. Note that the coefficients of skewness and kurtosis are not asked for, but are easily calculated as $\mu_3/\mu^{3/2}$ and $\mu_4/\mu_2^2 - 3$, respectively.

(a) For the Poisson distribution, it is typically easier to work with the *factorial moments*, but computing (central) moments remains a valid strategy. Let $(x)_m = \prod_{i=0}^{m}(x-i)$ denote the *falling factorial of degree m*. The k-th factorial moment is given by

$$E[(X)_k] = E\left[\prod_{i=0}^{k}(X-i)\right].$$

We can prove the following useful result for moments of X when $X \sim \mathcal{P}(\lambda)$:

Lemma (Poisson Factorial Moments). Let $X \sim \mathcal{P}(\lambda)$. Then for $k \in \{0, 1, 2, \ldots\}$

$$E[(X)_k] = \lambda^{k+1}.$$

Proof. Define $y = x - k$. Then

$$\begin{aligned}
E[(X)_k] &= \sum_{x=0}^{\infty} x(x-1)\cdots(x-k)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda^k \sum_{y=0}^{\infty} y\frac{e^{-\lambda}\lambda^y}{y!} \\
&= \lambda^{k+1}
\end{aligned}$$

■

This lets us write the raw moments by adding the appropriate factorial moments together.

$$\begin{aligned}
E(X) &= \lambda; \\
E[X(X-1)] &= E(X^2) - E(X) = \lambda^2; \\
E(X^2) &= \lambda^2 + \lambda; \\
E[X(X-1)(X-2)] &= E(X^3) - 3E(X^2) + 2E(X) = \lambda^3; \\
E(X^3) &= \lambda^3 + 3\lambda^2 + \lambda; \\
E[X(X-1)(X-2)(X-3)] &= E(X^4) - 6E(X^3) + 11E(X^2) - 6E(X) = \lambda^4; \\
E(X^4) &= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.
\end{aligned}$$

Now we need to use the raw moments in order to calculate the central moments.

$$\begin{aligned}
E(X) &= \lambda; \\
E(X-\lambda)^2 &= E(X^2) - \lambda^2 \\
&= \lambda; \\
E(X-\lambda)^3 &= E(X^3) - 3\lambda E(X^2) + 2\lambda^3 \\
&= \lambda; \\
E(X-\lambda)^4 &= E(X^4) - 4\lambda E(X^3) + 6\lambda^2 E(X^2) - 3\lambda^4 \\
&= 3\lambda^2 + \lambda.
\end{aligned}$$

(b) For the exponential distribution, either direct calculation of the (central) moments or use of the moment generating function are acceptable approaches. We demonstrate the mgf strategy in the form of the following lemma.

Lemma (Exponential Moments). If $X \sim \mathcal{E}(\lambda)$ (hazard parameterization, with mean $\mu = \lambda^{-1}$), then

$$E(X^m) = m!\lambda^{-m} = (EX)^m = m!\mu^m.$$

Proof. The mgf for X is

$$\begin{aligned} M_X(t) &= E(\exp(tX)) \\ &= \int_0^\infty \exp(tx)\lambda\exp(-\lambda x)dx \\ &= \lambda\int_0^\infty \exp(-(\lambda - t)x)dx \\ &= \frac{\lambda}{\lambda - t}, \quad t < \lambda. \end{aligned}$$

To obtain the m-th moment, we evaluate the m-th derivative of $M_X(t)$ at $t = 0$:

$$\frac{d^m}{dt^m}\frac{\lambda}{\lambda - t}\bigg|_{t=0} = \frac{m!\lambda}{(\lambda - t)^{m+1}}\bigg|_{t=0} = \frac{m!}{\lambda^m} = m!\mu^m.$$

Using this formula, we note that

$$\begin{aligned} EX^1 &= \mu; \\ EX^2 &= 2\mu^2; \\ EX^3 &= 6\mu^3; \\ EX^4 &= 24\mu^4. \end{aligned}$$

■

Hence, we have that the variance is given by

$$E(X - \mu)^2 = E(X^2) - \mu^2 = 2\mu^2 - \mu^2 = \mu^2.$$

We have also that the third central moment is

$$E(X - \mu)^3 = E(X^3) - \mu^3 - 3\mu E(X^2) + 3\mu^2 EX = 6\mu^3 - \mu^3 - 6\mu^3 + 3\mu^3 = 2\mu^3.$$

Finally, the fourth central moment is

$$E(X - \mu)^4 = E(X^4) + \mu^4 - 4\mu E(X^3) + 6\mu^2 E(X^2) - 4\mu^3 EX = 9\mu^4.$$

(c) For the Gaussian, there are several valid solutions to this problem. Either direct evaluation of the first four (central) moments or taking derivatives (at $t = 0$) of the moment generating function are both correct strategies. An additional approach to evaluating normal moments can be found in Stein's lemma.

Lemma (Stein). If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $g(x)$ is a differentiable function, then

$$E\left[(X - \mu)g(X)\right] = \sigma^2 Eg'(X).$$

Proof. The result follows by application of integration by parts to the left-hand side of the equality.

■

Now, we specialize Stein's lemma by choosing $g(x) = (x - \mu)^{k-1}$ for $k = 1, 2, 3, \ldots$ so that for $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$E(X - \mu)^k = (k - 1)\sigma^2 E(X - \mu)^{k-2}.$$

For $k = 1$, this gives us that $E(X - \mu) = 0$ so that

$$EX = \mu.$$

Now, for $k = 2$, this yields

$$E(X-\mu)^2 = \sigma^2.$$

For $k = 3$, we have

$$E(X-\mu)^3 = 2\sigma^2 E(X-\mu) = 0.$$

For $k = 4$, then

$$E(X-\mu)^4 = 3\sigma^2 E(X-\mu)^2 = 3\sigma^4.$$

3. Let X_1, X_2 be independent, identically distributed random variables. Find the density for random variable Y for the following combinations of distributions and transformations.

   (a) X_i has an exponential distribution with mean μ, and $Y = X_1^2 + X_2^2$.

   (b) $X_i \sim \mathbb{N}(\mu, \sigma^2)$, and $Y = X_1^2 + X_2^2$.

   (c) $X_i \sim \mathbb{U}(0, \theta)$, and $Y = \log(X_1 + X_2)$.

Ans:

(a) Let $\lambda = \dfrac{1}{\mu}$. Notice that on $[0, \infty)$, x^2 is a monotonous and smooth function. Hence, we can get the density of X_i^2 by applying the formula for transformed random variable. Let $W = X^2$

$$f_W(w) = \frac{\lambda}{2\sqrt{w}} e^{-\lambda\sqrt{w}}$$

which is density of a Weibull random variable with shape parameter 1/2 and scale parameter $1/\lambda^2$ or μ^2. Hence,

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_W(x) f_W(y-x)dx \\ &= \frac{\lambda^2}{4}\int_0^y \frac{e^{-\lambda(\sqrt{x}+\sqrt{y-x})}}{\sqrt{x(y-x)}}dx \\ &= \frac{1}{4\mu^2}\int_0^y \frac{e^{-(\sqrt{x}+\sqrt{y-x})/\mu}}{\sqrt{x(y-x)}}dx \end{aligned}$$

(b) Let $W = X^2$.

$$\begin{aligned} F_W(w) &= P(X^2 \le w) \\ &= P(-\sqrt{w} \le X \le \sqrt{w}) \\ &= \frac{1}{\sigma}\int_{-\sqrt{w}}^{\sqrt{w}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dw \\ \Rightarrow f_W(w) &= \frac{1}{\sigma}\left(\frac{1}{2\sqrt{w}} e^{-\dfrac{(\sqrt{w}-\mu)^2}{2\sigma^2}} + \frac{1}{2\sqrt{w}} e^{-\dfrac{(\sqrt{w}+\mu)^2}{2\sigma^2}}\right). \end{aligned}$$

To differentiate $F_W(w)$ w.r.t.w, we used Leibniz integral rule. The density can be re-written as

$$f_W(w) = \frac{1}{2\sigma\sqrt{w}} e^{-\dfrac{(\sqrt{w}-\mu)^2}{2\sigma^2}} \left(1 + e^{-\dfrac{2\mu\sqrt{w}}{\sigma^2}}\right).$$

When $\mu = 0$, the above becomes the density of a central chi-square. This is the density of a non-central Chi-square with degree of freedom 1, non-centrality parameter $\lambda = \mu^2$ and variance σ^2. In general, if X_i-s are i.i.d. $N(\mu_i, \sigma^2)$, then $\sum_{i=1}^n X_i^2$ is a non-central Chi-square with d.f. n, non-centrality parameter $\lambda = \sum_{i=1}^n \mu_i^2$, and variance σ^2. Therefore, we see that $X_1^2 + X_2^2$ in our case is also going to be a non-central Chi-square with d.f. 2, variance σ^2 and $\lambda = 2\mu^2$. Central Chi-squares are a special case of non-central Chi-squares when $\lambda = 0$. The density of non-central Chi-squares do not have a simple form. However, they can be represented as a mixture of central Chi-square and Poisson random variables. It can be proved that

$$\begin{aligned} &Y|J \sim \sigma^2\chi^2_{k+2J} \\ &J \sim Poi(\lambda/2) \\ &\Rightarrow Z \sim \text{ Non-central Chi-square with d.f. k, variance } \sigma^2 \text{ and noncentrality parameter } \lambda \end{aligned}$$

Using this representation and putting $k = 2$ and $\lambda = \mu^2$, the density f_Y becomes

$$f_Y(y) = \sum_{i=0}^{\infty} e^{-\mu^2} \frac{\mu^{2i}}{i!} \frac{1}{\sigma^2} \frac{\left(\frac{y}{\sigma^2}\right)^{(2+2i)/2-1} e^{-y/2\sigma^2}}{2^{(2+2i)/2}\Gamma(\frac{2+2i}{2})}$$

$$= \frac{1}{\sigma^2} \sum_{i=0}^{\infty} \frac{e^{-\frac{y/\sigma^2+\mu^2}{2}} \left(\frac{y}{\sigma^2}\right)^i \mu^{2i}}{(i!)^2 2^{i+1}}$$

However, for this homework it is sufficient to use the convolution formula to get the density of our Y.

$$f_Y(y) = \int_0^y f_W(x) f_W(y-x)dx$$

$$= \int_0^y \frac{\exp\left(-\frac{(\sqrt{x}-\mu)^2 + (\sqrt{y-x}-\mu)^2}{2\sigma^2}\right)}{(2\sigma)^2\sqrt{x(y-x)}} \left(1 + e^{-\frac{2\mu\sqrt{x}}{\sigma^2}}\right)\left(1 + e^{-\frac{2\mu\sqrt{y-x}}{\sigma^2}}\right)dx.$$

(c) Let $Z = X_1 + X_2$. Consider $z \in (0, \theta]$.

$$f_Z(z) = \int_0^z f(x)f(z-x)dx$$
$$= \int_0^z \frac{1}{\theta}\frac{1}{\theta}dx$$
$$= \frac{z}{\theta^2}.$$

Let $z \in (\theta, 2\theta]$. To have $z - x \le \theta$, we must have $x \ge z - \theta$. Hence, $f(z-x) = 0$ for $x \le z - \theta$.

$$f_Z(z) = \int_0^\theta f(x)f(z-x)dx$$
$$= \int_{z-\theta}^\theta \frac{1}{\theta}\frac{1}{\theta}dx$$
$$= \frac{2\theta - z}{\theta^2}.$$

Clearly for $z \in (2\theta, \infty)$, $f_Z(z) = 0$. $Y = \log Z$ and log is monotonic function. Therefore,

$$f_Y(y) = \begin{cases} \frac{e^{2y}}{\theta^2} & -\infty < y \le \log\theta \\ e^y \frac{2\theta - e^y}{\theta^2} & \log\theta < y \le \log 2\theta \\ 0 & o.w. \end{cases}$$

4. For each of the hierarchical models, find the density, mean, and variance of Y.

   (a) $X \sim \mathcal{N}(\mu, \tau^2)$ and $Y|X = x \sim \mathcal{N}(x, \sigma^2)$.

   (b) $X \sim \mathcal{U}(0, 1)$ and $Y|X = x \sim \mathcal{B}(n, x)$.

   (c) $X \sim \mathcal{P}(\lambda)$ and $Y|X = x \sim \mathcal{B}(x, p)$.

   (d) $X \sim \mathcal{E}(\lambda)$ and $Y|X = x \sim \mathcal{E}(x)$ (use the hazard parameterization)

Ans:

(a) We will denote by C a generic constant which may change value from line to line.

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} \frac{e^{-\left[\frac{(y-x)^2}{2\sigma^2}+\frac{(x-\nu)^2}{2\tau^2}\right]}}{2\pi\sigma\tau}dx \\
&= C\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\frac{\tau^2+\sigma^2}{\tau^2\sigma^2}\left[x^2-2x\left(\frac{y\tau^2+\nu\sigma^2}{\tau^2+\sigma^2}\right)+\left(\frac{y\tau^2+\nu\sigma^2}{\tau^2+\sigma^2}\right)^2\right]} e^{-\frac{y^2}{2\sigma^2}+\frac{1}{2}\left(\frac{y\tau^2+\nu\sigma^2}{\tau^2+\sigma^2}\right)^2\frac{\tau^2+\sigma^2}{\tau^2\sigma^2}}}{2\pi\sigma\tau}dx \\
&= Ce^{-\frac{1}{2}\frac{y^2}{\sigma^2}+\frac{y^2\tau^4+2y\nu\sigma^2\tau^2+\sigma^4\nu^2}{2\sigma^2\tau^2(\sigma^2+\tau^2)}}
\end{aligned}$$

since the rest of the integrand is a constant times a normal density, it will integrate to a constant.

$$\begin{aligned}
f_Y(y) &= Ce^{-\frac{1}{2}\frac{y^2-2y\nu}{\sigma^2+\tau^2}} \\
&= Ce^{-\frac{1}{2}\frac{(y-\nu)^2}{\sigma^2+\tau^2}}
\end{aligned}$$

Since $f_Y(y) \sim e^{-\frac{1}{2}\frac{(y-\nu)^2}{\sigma^2+\tau^2}}$, $Y \sim \mathcal{N}(\mu, \sigma^2+\tau^2)$ since the normalizing constant is forced to be $\frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}$ as $f_Y(y)$ integrates to 1. Notice that in this example, we did not explicitly calculate the constant multipliers in each step. It is sufficient to identify the terms involving y only.

$EY = \nu,\ Var(Y) = \sigma^2+\tau^2.$

(b)

$$\begin{aligned}
P(Y=k) &= \binom{n}{k}\int_0^1 x^k(1-x)^{n-k}dx \\
&= \binom{n}{k}B(k+1,n-k+1)\int_0^1 \frac{x^{k+1-1}(1-x)^{n-k+1-1}}{B(k+1,n-k+1)}dx \\
&= \frac{n!}{(n-k)!k!}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n-k+1+k+1)} \\
&= \frac{1}{n+1}
\end{aligned}$$

Here we used the fact that the density of $Beta(k+1,n-k+1)$ is $f(x) = \frac{x^{k+1-1}(1-x)^{n-k+1-1}}{B(k+1,n-k+1)}$ for $x \in (0,1)$. We notice that $Y \sim \mathcal{U}\{0,n\}$. Hence,

$$\begin{aligned}
EY &= \frac{n+0}{2} = \frac{n}{2} \\
Var(Y) &= \frac{(n-0+1)^2-1}{12} = \frac{n^2+2n}{12}.
\end{aligned}$$

(c)

$$\begin{aligned}
P(Y=k) &= \sum_{x=k}^{\infty}\binom{x}{k}p^k(1-p)^{x-k}\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=k}^{\infty}\frac{p^k(1-p)^{x-k}e^{-\lambda}\lambda^x}{k!(x-k)!} \\
&= \sum_{r=0}^{\infty}\frac{e^{-\lambda}(\lambda p)^k}{k!}\frac{(\lambda(1-p))^r}{r!}
\end{aligned}$$

where $r = x-k$.

$$\begin{aligned}
P(Y=k) &= \frac{e^{-\lambda}(\lambda p)^k}{k!}e^{\lambda(1-p)} \\
&= \frac{e^{-\lambda p}(\lambda p)^k}{k!}
\end{aligned}$$

Hence, $Y \sim \mathcal{P}(\lambda p)$. Therefore

$$\begin{aligned} EY &= \lambda p \\ Var(Y) &= \lambda p. \end{aligned}$$

(d)

$$\begin{aligned} f_Y(y) =& \int_0^\infty xe^{-xy}e^{-\lambda x}dx \\ =& -x\frac{e^{-(\lambda+y)x}}{(\lambda+y)}\bigg|_0^\infty + \int_0^\infty \frac{e^{-(\lambda+y)x}}{\lambda+y}dx \\ =& 0 - 0 - \frac{e^{-(\lambda+y)x}}{(\lambda+y)^2}\bigg|_0^\infty \\ =& \frac{1}{(\lambda+y)^2} \end{aligned}$$

$$\begin{aligned} EY =& \int_0^\infty \frac{y}{(\lambda+y)^2}dy \\ =& \int_0^\infty \frac{\lambda+y}{(\lambda+y)^2}dy - \lambda\int_0^\infty \frac{dy}{(\lambda+y)^2} \\ =& \log(\lambda+y)\bigg|_0^\infty + \frac{\lambda}{\lambda+y}\bigg|_0^\infty \end{aligned}$$

The first term is ∞ where the second term is finite. Hence, $EY = \infty$. Therefore, $EY^2 \geq (EY)^2 = \infty$. Therefore, $Var(Y) = \infty$.

MORE INVOLVED PROBLEMS

5. We consider a sequential experiment in which we have potential observations X_1 and X_2 which are independent and identically distributed $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Our sequential sampling plan is as follows: We observe X_1, and if, for some prespecified $a < b$, $X_1 \leq a$ or $X_1 \geq b$, we stop. Otherwise we continue sampling to observe X_2. At the end of our experiment, we have the bivariate sequential test statistic

$$(M, S) = \begin{cases} (1, X_1) & X_1 \leq a \text{ or } X_1 \geq b \\ (2, X_1 + X_2) & \text{otherwise} \end{cases}$$

In order to estimate the unknown mean μ, we use the observed sample mean $\hat{\mu} = S/M$.

(a) Find the density for (M, S). (This cannot be solved in closed form, so it is sufficient to write down the integral you would use to find it.)

(b) Suppose $\mu = 0, \sigma^2 = 1, a = 0, b = 2.7897$.

i. Find $Pr[M = 1, S \leq a]$.

ii. Find $Pr[M = 1, S \geq b]$.

iii. Find $Pr[M = 2, S \leq 0]$.

iv. Find the value of c such that $Pr[M = 1, S \geq b] + Pr[M = 2, S \geq c] = 0.025$.

(c) Derive a formula for the expected value for $\hat{\mu}$ in the general case. Under what conditions will $E[\hat{\mu}] = \mu$? If an estimator $\hat{\mu}$ satisfies $E[\hat{\mu}] = \mu, \forall \mu$, we call that estimator unbiased. Under what conditions on our sampling plan is $\hat{\mu}$ unbiased?

(a) We consider separately the cases for $M = 1$ and $M = 2$. First

$$\begin{aligned} Pr[M = 1, S \leq s] &= Pr[X_1 \in \{(-\infty, a] \cup [b, \infty), X_1 \leq s] \\ &= \int_{-\infty}^{s} \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathbf{1}[x \notin (a, b)]dx, \end{aligned}$$

where ϕ is the density for the standard normal distribution, $\mathcal{N}(0, 1)$. Differentiating with respect to s, we obtain the sub-density for $M = 1$:

$$p_1(s) = \frac{1}{\sigma}\phi\left(\frac{s-\mu}{\sigma}\right) \mathbf{1}[s \notin (a, b)].$$

Now, the second case is

$$\begin{aligned} Pr[M = 2, S \leq s] &= Pr[X_1 \in (a, b), S \leq s] \\ &= \int_{-\infty}^{s} \int_{(a,b)} \frac{1}{\sigma}\phi\left(\frac{y-x-\mu}{\sigma}\right) \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) dxdy. \end{aligned}$$

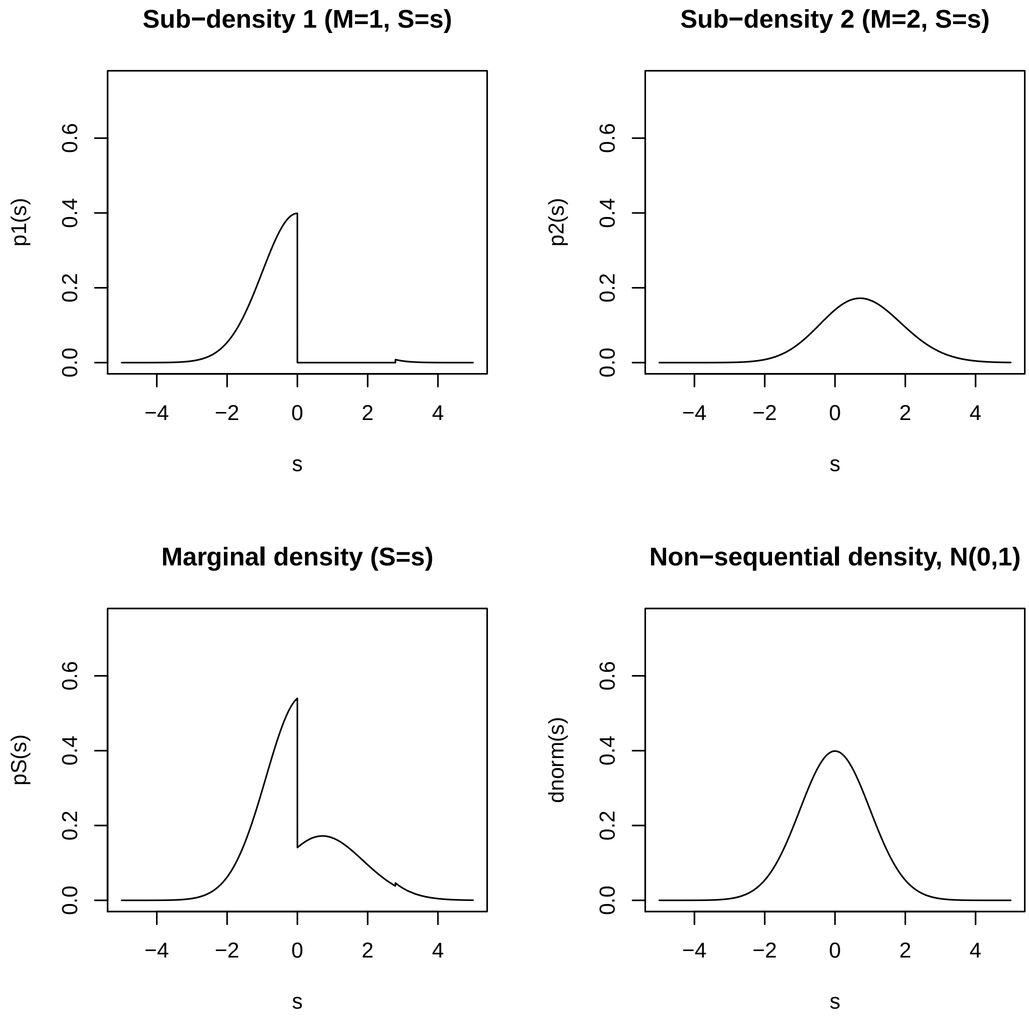
Differentiating with respect to s, we optain the sub-density for $M = 2$:

$$p_2(s) = \int_{(a,b)} \frac{1}{\sigma}\phi\left(\frac{s-x-\mu}{\sigma}\right) \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) dx.$$

Hence, the joint density is given by

$$p(m, s) = \begin{cases} p_1(s) & m = 1, s \notin (a, b) \\ p_2(s) & m = 2 \\ 0 & \text{else.} \end{cases}$$

We plot the sub-densities below, alongside the marginal density for S and for a standard normal distribution. Notice that the sub-densities do not integrate to 1.



(b) We numerically integrate $p_1(s)$ and $p_2(s)$ to obtain the following probabilities. Depending on how your implemented your numeric algorithm, there may be slight differences in your answers.

i. $Pr[M = 1, S \leq a] = 0.5$.

ii. $Pr[M = 1, S \geq b] = 0.002637847$.

iii. $Pr[M = 2, S \leq 0] = 0.1249965$.

iv. The value of c such that $Pr[M = 1, S \geq b] + Pr[M = 2, S \geq c] = 0.025$ is $c = 2.789718$.

(c) For $\hat{\mu} = S/M$, we have the following expectation.

$$
\begin{aligned}
E(\hat{\mu}) &= \sum_{m=1}^{2} \int_{-\infty}^{\infty} \frac{s}{m} p_m(s) ds \\
&= \int_{-\infty}^{a} s p_1(s) ds + \int_{b}^{\infty} s p_1(s) ds + \int_{-\infty}^{\infty} \frac{s}{2} p_2(s) ds \\
&= \underbrace{\int_{-\infty}^{\infty} s p_1(s) ds}_{=\mu} - \underbrace{\int_{a}^{b} s p_1(s) ds}_{\equiv E_1} + \underbrace{\int_{-\infty}^{\infty} \frac{s}{2} p_2(s) ds}_{\equiv E_2} \\
&= \mu - E_1 + E_2
\end{aligned}
$$

We find that E_1 is given by

$$\begin{aligned}
E_1 &= \int_a^b s p_1(s) ds \\
&= \int_a^b (s-\mu) p_1(s) ds + \int_a^b \mu p_1(s) ds \\
&= \int_a^b (s-\mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds + \mu \int_a^b \frac{1}{\sigma} \phi\left(\frac{s-\mu}{\sigma}\right) ds \\
&= \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\frac{(a-\mu)^2}{2\sigma^2}} - e^{-\frac{(b-\mu)^2}{2\sigma^2}} \right) + \mu \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right).
\end{aligned}$$

Now, for E_2 we have

$$\begin{aligned}
E_2 &= \int_{-\infty}^{\infty} \frac{s}{2} p_2(s) ds \\
&= \int_{-\infty}^{\infty} \frac{s}{2} \int_{(a,b)} \frac{1}{\sigma} \phi\left(\frac{s-x-\mu}{\sigma}\right) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx ds \\
&= \int_{(a,b)} \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \underbrace{\int_{-\infty}^{\infty} \frac{s}{2} \frac{1}{\sigma} \phi\left(\frac{s-x-\mu}{\sigma}\right) ds}_{(x+\mu)/2} dx \\
&= \int_{(a,b)} \frac{(x+\mu)}{2\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx \\
&= \frac{\sigma}{2\sqrt{2\pi}} \left(e^{-\frac{(a-\mu)^2}{2\sigma^2}} - e^{-\frac{(b-\mu)^2}{2\sigma^2}} \right) + \mu \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right).
\end{aligned}$$

All together, we have that

$$E(\hat{\mu}) = \mu - E_1 + E_2 = \mu - \frac{\sigma}{2\sqrt{2\pi}} \left(e^{-\frac{(a-\mu)^2}{2\sigma^2}} - e^{-\frac{(b-\mu)^2}{2\sigma^2}} \right).$$

Hence, $\hat{\mu}$ is unbiased if and only if $a < b$ are such that $\mu = (a+b)/2$.

6. Consider a sample of independent, identically distributed continuous random variables $X_i \sim F_X$, $i = 1\ldots,n$, and a sample of independent, identically distributed continuous random variables $Y_j \sim 1\ldots,n$, where the X's and the Y's are also totally independent. We are interested in whether the X_i 's have the same distribution as the Y_i 's, and we choose consider estimating the probability θ that a randomly chosen X will be greater than a randomly chosen Y.
   (a) Provide a formula for θ in terms of the joint distribution of (X_i, Y_i).
   (b) Let $U_{i,j} = 1_{[X_i \geq Y_j]}$. Find the probability distribution for U_{ij}.
   (c) What is the mean and variance for U_{ij}?
   (d) What is the covariance $Cov(U_{ij}, U_{kl})$ for arbitrary i, j, k, l? (Be sure to consider cases when $i = k$ and, or $j = l$.)
   (e) Are the random variables $\{U_{ij} : 1 \leq i \leq n; 1 \leq j \leq m\}$ identically distributed? Are they totally independent?
   (f) Now consider the statistic

   $$U = \sum_{i=1}^{n} \sum_{j=1}^{m} 1_{[X_i \geq Y_j]}.$$

   Find the expectation and variance of U when $F_X(u) = F_Y(u), \forall u \in \mathbb{R}$.

Ans:

(a) since X and Y are independent,

$$\begin{aligned}
\theta &= P(X \geq Y) \\
&= \int_{-\infty}^{\infty} \int_{y}^{\infty} f_{X,Y}(x,y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{y}^{\infty} f_X(x) f_Y(y) dx dy.
\end{aligned}$$

(b) $U_{ij} \sim Ber(p)$ where $p = P(X_i \geq Y_j) = \theta$.

(c) Since $U_{ij} \sim Ber(\theta)$,

$$EU_{ij} = \theta; \; VarU_{ij} = \theta(1-\theta) \tag{1}$$

(d) Clearly when $i \neq k, j \neq l$, $Cov(U_{ij}, U_{kl}) = 0$ by the independence of the random variables X_i and Y_j-s.

Also, when $i = k, j = l$, $Cov(U_{ij}, U_{kl}) = Cov(U_{ij}, U_{ij}) = Var(U_{ij}) = \theta(1-\theta)$. Hence, the nontrivial cases are:

1. $i = k, j \neq l$

$$\begin{aligned}
&Cov(U_{ij}, U_{il}) \\
&= EU_{ij}U_{il} - EU_{ij}EU_{il} \\
&= P(X_i \geq y_j, X_i \geq Y_l) - \theta^2 \\
&= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{x} f_Y(y_j) \int_{-\infty}^{x} f_Y(y_l) dy_l dy_j dx - \theta^2 \\
&= \int_{-\infty}^{\infty} f_X(x) F_Y(x)^2 dx - \theta^2 \neq 0,
\end{aligned}$$

in general. Also notice that when $F_X = F_y = F$ with pdf f,

$$\begin{aligned}
&\int_{-\infty}^{\infty} f_X(x) F_Y(x)^2 dx - \theta^2 \\
&= \int_{-\infty}^{\infty} f(x) F(x)^2 dx - \frac{1}{4} \\
&= \int_{-\infty}^{\infty} F(x)^2 dF(x) - \frac{1}{4} \\
&= \int_{0}^{1} z^2 dz - \frac{1}{4} \\
&= \frac{1}{12}
\end{aligned}$$

Hence when $F_X = F_y = F$,

$$Cov(U_{ij}, U_{il}) = \frac{1}{12}. \tag{2}$$

2. $i \neq k, j = l$

$$\begin{aligned}
&Cov(U_{ij}, U_{kj}) \\
&= EU_{ij}U_{kj} - EU_{ij}EU_{kj} \\
&= P(X_i \geq y_j, X_k \geq Y_l) - \theta^2 \\
&= \int_{-\infty}^{\infty} f_Y(y) \int_{y}^{\infty} f_X(x_i) \int_{y}^{\infty} f_X(x_k) dx_k dx_i dy - \theta^2 \\
&= \int_{-\infty}^{\infty} (1 - F_X(y))^2 f_Y(y) dy - \theta^2 \neq 0,
\end{aligned}$$

in general. Also notice that when $F_X = F_Y = F$ with pdf f,

$$\begin{aligned}
&\int_{-\infty}^{\infty} (1 - F_X(y))^2 f_Y(y) dy - \theta^2 \\
&= \int_{-\infty}^{\infty} (1 - F(y))^2 f(y) dy - \frac{1}{4} \\
&= \int_{-\infty}^{\infty} (1 - F(y))^2 dF(y) - \frac{1}{4} \\
&= \int_{0}^{1} (1 - z)^2 dz - \frac{1}{4} \\
&= \frac{1}{12}
\end{aligned}$$

Hence when $F_X = F_y = F$,

$$Cov(U_{ij}, U_{kj}) = \frac{1}{12}. \tag{3}$$

(e) Since $U_{ij} \sim Ber(\theta) \forall i, j$, they are identicaly distributed. They may not always be independent since their covariances may not equal to 0 when $i = k$ or $j = l$.

(f) When $F_X = F_Y,\ \theta = \frac{1}{2}$. Hence, it is easy to see that

$$EU = mn\theta = \frac{mn}{2}$$

$$\begin{aligned}
Var(U) &= \sum_{i=1}^{n}\sum_{j=1}^{m} Var(U_{ij}) + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=j+1}^{m} Cov(U_{ij}, U_{il}) + 2\sum_{i=1}^{n}\sum_{k=i+1}^{n}\sum_{j=1}^{m} Cov(U_{ij}, U_{kj}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \theta(1-\theta) + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=j+1}^{m} (\int_{-\infty}^{\infty} f_X(x)F_Y(x)^2 dx - \theta^2) \\
&+ 2\sum_{i=1}^{n}\sum_{k=i+1}^{n}\sum_{j=1}^{m} (\int_{-\infty}^{\infty} (1 - F_X(y))^2 f_Y(y) dy - \theta^2) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \theta(1-\theta) + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=j+1}^{m} \frac{1}{12} + 2\sum_{i=1}^{n}\sum_{k=i+1}^{n}\sum_{j=1}^{m} \frac{1}{12} \text{ (Using (3) and (2))} \\
&= \frac{mn(1+m+n)}{12}
\end{aligned}$$

7. Consider again the setting of the previous problem, but that we transform all of the random variables from the scale they were originally measured on to their ranks in the combined sample

$$\begin{aligned}
R_i = rank(X_i) &= \sum_{j=1}^{n} 1_{[X_j \le X_i]} + \sum_{j=1}^{m} 1_{[Y_j \le X_i]} \\
S_i = rank(Y_i) &= \sum_{j=1}^{n} 1_{[X_j \le Y_i]} + \sum_{j=1}^{m} 1_{[Y_j \le Y_i]}
\end{aligned}$$

and define

$$R = \sum_{i=1}^{n} R_i$$

(a) Find the mean and variance of R when $F_X(u) = F_Y(u), \forall u \in \mathbb{R}$. (Hint: Under the null hypothesis, R has the same distribution as the sum of m numbers randomly chosen without replacement from the integers $\{1, \ldots, n\}$.)

(b) Find the correlation between U from the previous problem and R as defined in this problem.

Ans:

(a) When $F_X(u) = F_Y(u), \forall u \in \mathbb{R}$, it is like X_i is being sampled from a sample of $m + n$ i.i.d. random variables. Therefore, each rank from 1 to $m + n$ is equally likely for R_i. Hence, $R_i \sim \mathcal{U}\{1, m+n\}$. Therefore,

$$\begin{aligned}
ER_i &= \sum_{i=1}^{m+n} \frac{i}{m+n} = \frac{m+n+1}{2}. \\
ER &= \frac{n(m+n+1)}{2}.
\end{aligned}$$

Notice that

$$\begin{aligned}
R &= \sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_i]} + \sum_{i=1}^{n}\sum_{j=1}^{m} 1_{[Y_j \le X_i]} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_i]} + U
\end{aligned}$$

Also, $\sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_i]} = \sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_{(i)}]}$ where $\{X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}\}$ is the order statistic. therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_i]} = \sum_{i=1}^{n}\sum_{j=1}^{n} 1_{[X_j \le X_{(i)}]} = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Therefore,

$$R = U + \frac{n(n+1)}{2} \tag{4}$$

Hence, $Var(R) = Var(U) = \dfrac{mn(1+n+m)}{12}$ by 6.(f).

(b) We obtained that

$$R = U + \frac{n(n+1)}{2}.$$

Hence,

$$\begin{aligned} &Cov(R,U) = Var(U) \\ &Var(R) = Var(U) \\ &Cor(R,U) = 1. \end{aligned}$$

APPENDIX: R-code for Question 5

```
## Illustrating 5(a)

# The density component when m=1
p1 <- function(s, mean=0, sd=1, a=0, b=2.7897){
  dnorm(s, mean, sd)*(1-(a<s)*(s<b))
}

# The density component when m=2
p2 <- Vectorize( function(s, mean=0, sd=1, a=0, b=2.7897){
  int <- function(x){ dnorm(s-x, mean, sd)*dnorm(x, mean, sd) }
  integrate(int, a, b)$value
}, "s")

# The marginal density for S
pS <- function(s, mean=0, sd=1, a=0, b=2.7897){
  p1(s, mean, sd, a, b) + p2(s, mean, sd, a, b)
}

# The marginal density for M
pM <- function(m, mean=0, sd=1, a=0, b=2.7897){
  f1 <- function(s) p1(s, mean, sd, a, b)
  f2 <- function(s) p2(s, mean, sd, a, b)
  ifelse(m==1, integrate(f1, -Inf, Inf)$value, integrate(f2, -Inf, Inf)$value )
}

# View each of the densities
pdf("Q5plot.pdf")
par(mfrow=c(2,2))
curve(p1, -5, 5, n=1e4, xname="s", ylim=0:1*.75, main="Sub-density 1 (M=1, S=s)")
curve(p2, -5, 5, n=1e4, xname="s", ylim=0:1*.75, main="Sub-density 2 (M=2, S=s)")
curve(pS, -5, 5, n=1e4, xname="s", ylim=0:1*.75, main="Marginal density (S=s)")
curve(dnorm, -5, 5, n=1e4, xname="s", ylim=0:1*.75, main="Non-sequential density, N(0,1)")
dev.off()

## Part 5(b)(i-iii)

# P(M=1, S<=0)
b.i <- integrate(p1, -Inf, 0)$value

# P(M=1, S>=2.7897)
b.ii <- integrate(p1, 2.7897, Inf)$value

# P(M=2, S<=0)
b.iii <- integrate(p2, -Inf, 0)$value

## Part 5(b)(iv)
objective <- function(x) integrate(p2, x, Inf)$value + b.ii - 0.025
c <- uniroot(objective, c(0,3))$root
```