Homework 4

ALECK ZHAO

October 4, 2016

1. Consider a finite population of size N of objects, and suppose for each object i there are a pair of measurements (x_i, y_i). Suppose we obtain a sample of size n of these pairs of values from this population, where we sample without replacement. Let $\bar{X}$ and $\bar{Y}$ be the sample means of the x-measurements and the y-measurements, respectively. We claimed in lecture that

$$\text{Cov}(\bar{X}, \bar{Y}) = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right). \tag{1}$$

   Here σ_{xy} is the population covariance, defined by

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

   where μ_x and μ_y are the population means, respectively, of the x- and y-measurements. Prove that Equation 1 holds.

   Proof. We have

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i, \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n}Y_i$$

   where X_i and Y_i are measurements of the sample. Then the covariance is given by

$$\begin{aligned}
\text{Cov}(\bar{X}, \bar{Y}) &= \text{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}X_i, \frac{1}{n}\sum_{j=1}^{n}Y_j\right) \\
&= \frac{1}{n^2}\text{Cov}\left(\sum_{i=1}^{n}X_i, \sum_{j=1}^{n}Y_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(X_i, Y_j)
\end{aligned}$$

   Here, if $i = j$, then $\text{Cov}(X_i, Y_j) = \text{Cov}(X_i, Y_i) = \sigma_{xy}$ is the population covariance. Otherwise let $i \neq j$ and we have

$$\text{Cov}(X_i, Y_j) = E[X_iY_j] - E[X_i][Y_j] = E[X_iY_j] - \mu_x\mu_y.$$

   Now, considering X_iY_j as a random variable, we have

$$E[X_iY_j] = \sum xy \cdot P(X_i = x, Y_j = y),$$

   where the probability is the joint distribution. Suppose the distinct values of x_i and y_i be represented by the sets

$$\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$$
$$\{\beta_1, \beta_2, \cdots, \beta_\ell\}$$

respectively, where $k, \ell \leq n$. Then suppose n_a and m_b represent the number of times α_a and β_b appear in the sample, respectively. Now, rewrite the expectation as

$$E[X_i Y_j] = \sum_{a=1}^{k} \sum_{b=1}^{\ell} \alpha_a \beta_b \cdot P(X_i = \alpha_a, Y_j = \beta_b)$$

$$= \sum_{a=1}^{k} \sum_{b=1}^{\ell} \alpha_a \beta_b P(X_i = \alpha_a | Y_j = \beta_b) P(Y_j = \beta_b)$$

$$= \sum_{b=1}^{\ell} \left(\beta_b P(Y_j = \beta_b) \cdot \sum_{a=1}^{k} \alpha_a P(X_i = \alpha_a | Y_j = \beta_b) \right)$$

$$= \sum_{b=1}^{\ell} \left(\beta_b \frac{m_b}{N} \cdot \sum_{a=1}^{k} \alpha_a P(X_i = \alpha_a | Y_j = \beta_b) \right)$$

The right summation split into two parts. If (α_a, β_b) belong to the same pair (x_i, y_i), then if $Y_j = \beta_b$, since $i \neq j$, there are $n_a - 1$ remaining elements of x with value α_a, out of a total of $N - 1$ remaining elements total, so the probability is $\frac{n_a - 1}{N-1}$.

If (α_a, β_b) don't belong to the same pair (x_i, y_i), then there are still n_a remaining α_a values left, and $N - 1$ total, so the probability is $\frac{n_a}{N-1}$.

The case when (α_a, β_b) belong to the same pair only occurs once in the summation, suppose when $a = p(b)$ where p is a function of b, so it can split up as

$$\sum_{a=1}^{k} \alpha_a P(X_i = \alpha_a | Y_j = \beta_b) = \sum_{a=p} \alpha_a P(X_i = \alpha_a | Y_j = \beta_b) + \sum_{a \neq p} \alpha_a P(X_i = \alpha_a | Y_j = \beta_b)$$

$$= \alpha_p \frac{n_p - 1}{N - 1} + \sum_{a \neq p} \alpha_a \frac{n_a}{N - 1}$$

$$= \alpha_p \frac{n_p}{N - 1} - \frac{\alpha_p}{N - 1} + \sum_{a \neq p} \alpha_a \frac{n_a}{N - 1}$$

$$= -\frac{\alpha_p}{N - 1} + \sum_{a=1}^{k} \alpha_a \frac{n_a}{N - 1}$$

$$= -\frac{\alpha_p}{N - 1} + \frac{\tau_x}{N - 1}$$

Now, substituting this into the summation above, we have

$$E[X_i Y_j] = \sum_{b=1}^{\ell} \left(\beta_b \frac{m_b}{N} \cdot \left(\frac{\tau_x - \alpha_p}{N - 1} \right) \right)$$

$$= \frac{1}{N(N - 1)} \sum_{b=1}^{\ell} (\beta_b m_b \tau_x - \beta_b m_b \alpha_{p(b)})$$

$$= \frac{1}{N(N - 1)} \left(\tau_x \sum_{b=1}^{\ell} \beta_b m_b - \beta_b m_b \alpha_{p(b)} \right)$$

$$= \frac{1}{N(N - 1)} \left(\tau_x \tau_y - \sum_{b=1}^{\ell} \beta_b m_b \alpha_{p(b)} \right)$$

$$= \frac{\tau_x \tau_y}{N(N - 1)} - \frac{1}{N(N - 1)} \sum_{i=1}^{N} x_i y_i$$

where the second summation was taking the sum of every instance where the two values $\alpha_{p(b)}$ and β_b belonged to the same pair, which is the same as summing over all pairs (x_i, y_i) in the population.

Manipulating the population variance, we have

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)$$

$$= \frac{1}{N} \left(\sum_{i=1}^{N} x_i y_y - \mu_x \sum_{i=1}^{N} y_i - \mu_y \sum_{i=1}^{N} x_k + \sum_{i=1}^{N} \mu_x \mu_y \right)$$

$$= \frac{1}{N} \left(\sum_{i=1}^{N} x_i y_i - \mu_x \tau_y - \mu_y \tau_x + N \mu_x \mu_y \right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i y_i - \mu_x \frac{\tau_y}{N} - \mu_y \frac{\tau_x}{N} + \mu_x \mu_y$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i y_i - \mu_x \mu_y$$

$$\implies N(\sigma_{xy} + \mu_x \mu_y) = \sum_{i=1}^{N} x_i y_i$$

Substituting this above, we have

$$E[X_i Y_j] = \frac{\tau_x \tau_y}{N(N-1)} - \frac{N(\sigma_{xy} + \mu_x \mu_y)}{N(N-1)}$$

$$= \frac{N^2 \mu_x \mu_y - N \sigma_{xy} - N \mu_x \mu_y}{N(N-1)}$$

$$= \frac{(N^2 - N)\mu_x \mu_y - N \sigma_{xy}}{N(N-1)}$$

$$= \mu_x \mu_y - \frac{\sigma_{xy}}{N-1}$$

$$\implies \mathrm{Cov}(X_i, Y_j) = E[X_i Y_j] - \mu_x \mu_y$$

$$= -\frac{\sigma_{xy}}{N-1}$$

Finally, using this in the summation for $\mathrm{Cov}(\bar{X}, \bar{Y})$, we have

$$\mathrm{Cov}(\bar{X}, \bar{Y}) = \frac{1}{n^2} \left(\sum_{i=j} \sigma_{xy} + \sum_{i \neq j} -\frac{\sigma_{xy}}{N-1} \right)$$

$$= \frac{1}{n^2} \left(n \sigma_{xy} + n(n-1) \left(-\frac{\sigma_{xy}}{N-1} \right) \right)$$

$$= \frac{\sigma_{xy}}{n} \left(1 - \frac{n-1}{N-1} \right)$$

as desired.

$\square$

2. Let $X_1, \cdots, X_n$ be an iid sample from some distribution having mean μ and variance σ^2. Recall that the CLT says that $\sqrt{n}(\bar{X} - \mu)$ has an approximate $N(0, \sigma^2)$ distribution as n tends to infinity. Let g be a smooth function with $g'(\mu) \neq 0$.

(a) Write down a first order Taylor's expansion with remainder for g about μ.

Solution. We have
$$g(x) \approx g(\mu) + g'(\mu)(x - \mu) + \frac{1}{2}g''(\mu)(x - \mu)^2 + R.$$

$\square$

(b) Substitute $\bar{X}$ for x in the expansion in (a) and use this to express $\sqrt{n}(g(\bar{X}) - g(\mu))$ as a sum of two terms, one involving $(\bar{X} - \mu)$ and the other involving $(\bar{X} - \mu)^2$.

Solution. We have
$$g(\bar{X}) \approx g(\mu) + g'(\mu)(\bar{X} - \mu) + \frac{1}{2}g''(\mu)(\bar{X} - \mu)^2 + R$$

so using this, we have
$$\sqrt{n}(g(\bar{X}) - g(\mu)) \approx \sqrt{n}\left[\left(g(\mu) + g'(\mu)(\bar{X} - \mu) + \frac{1}{2}g''(\mu)(\bar{X} - \mu)^2 + R\right) - g(\mu)\right]$$
$$= \sqrt{n}g'(\mu)(\bar{X} - \mu) + \frac{\sqrt{n}g''(\mu)}{2}(\bar{X} - \mu)^2 + R$$

as desired.

$\square$

(c) Explain why the linear term should have a distribution that is approximately $N(0, g'(\mu)^2\sigma^2)$ as n tends to infinity. (Hint: If $Y \sim N(0, \sigma^2)$ what is the distribution of cY where c is a constant?)

Solution. Since we know $Y = \sqrt{n}(\bar{X} - \mu)$ is approximately $N(0, \sigma^2)$, it follows that the distribution of $g'(\mu)Y \sim N(0, g'(\mu)^2\sigma^2)$, as desired.

$\square$

(d) Explain why the quadratic term should tend to zero as n tends to infinity.

Solution. By the Law of Large Numbers, the random variable $(\bar{X} - \mu)^2$ should approach its true value as $n \to \infty$, which is $E[(\bar{X} - \mu)^2] = \sigma^2/n$. Then
$$\frac{\sqrt{n}g''(\mu)}{2}\frac{\sigma^2}{n} = \frac{g''(\mu)}{2}\frac{\sigma^2}{\sqrt{n}} \to 0$$

as $n \to \infty$, as desired.

$\square$

3. Suppose X_i are iid Poisson random variables with parameter λ.

   (a) Determine the distribution of the sum $S = \sum_{i=1}^{N} X_i$. (Hint: consider the case $n = 2$ and then use induction.)

   Solution. We claim the distribution of the sum $S_n = \sum_{i=1}^{n} X_i$ is given by

   $$p_{S_n}(s) = \frac{(n\lambda)^s e^{-n\lambda}}{s!}$$

   which is just a Poisson with parameter $n\lambda$.

   Let $S_2 = X_1 + X_2$. Let $p(x)$ be the density of both X_1 and X_2. Then by convolution,

   $$\begin{aligned}
   p_{S_2}(s) &= \sum_{i=0}^{s} p(i)p(s-i) \\
   &= \sum_{i=0}^{s} \frac{\lambda^i e^{-\lambda}}{i!} \cdot \frac{\lambda^{s-i} e^{-\lambda}}{(s-i)!} \\
   &= \frac{\lambda^s e^{-2\lambda}}{s!} \sum_{i=0}^{s} \frac{s!}{i!(s-i)!} = \frac{\lambda^s e^{-2\lambda}}{s!} \sum_{i=0}^{s} \binom{s}{i} \\
   &= \frac{\lambda^s e^{-2\lambda}}{s!} 2^s = \frac{(2\lambda)^s e^{-2\lambda}}{s!}
   \end{aligned}$$

   so the base case is true.

   Now, suppose this closed form expression for the distribution holds for arbitrary k, so that

   $$p_{S_k}(s) = \frac{(k\lambda)^s e^{-k\lambda}}{s!}.$$

   Then the distribution of $S_{k+1} = S_k + X_{k+1}$ can be calculated by convolution:

   $$\begin{aligned}
   f_{S_{k+1}}(s) &= \sum_{i=0}^{s} p_{S_k}(i)f(s-i) \\
   &= \sum_{i=0}^{s} \frac{(k\lambda)^i e^{-k\lambda}}{i!} \cdot \frac{\lambda^{s-i} e^{-\lambda}}{(s-i)!} \\
   &= \frac{\lambda^s e^{-(k+1)\lambda}}{s!} \sum_{i=0}^{s} k^i \cdot \frac{s!}{i!(s-i)!} \\
   &= \frac{\lambda^s e^{-(k+1)\lambda}}{s!} \sum_{i=0}^{s} k^i \binom{s}{i} \\
   &= \frac{\lambda^s e^{-(k+1)\lambda}}{s!} (1+k)^s \\
   &= \frac{[(k+1)\lambda]^s e^{-(k+1)\lambda}}{s!} \\
   &= p_{S_{k+1}}(s)
   \end{aligned}$$

   so the claim is true, as desired.

   $\square$

(b) Obtain the method of moments estimator for λ.

Solution. We have

$$
\begin{aligned}
\mu_1 = E[X_i] &= \sum_{j=0}^{\infty} j \cdot \frac{\lambda^j e^{-\lambda}}{j!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda
\end{aligned}
$$

where the summation was the Taylor expansion for e^x at λ. Thus, the method of moments estimator for λ is given by

$$
\hat{\lambda} = \hat{\mu}_1 = \bar{X} = S/n.
$$

$\square$

(c) Assuming n is large, obtain an approximate 95% confidence interval for λ. Your confidence interval should be in terms of your confidence level, your estimator, n, etc.

Solution. We have

$$
E[\hat{\lambda}] = E\left[\frac{S}{n}\right] = \frac{1}{n}E[S] = \frac{1}{n}n\lambda = \lambda
$$

$$
\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{S}{n}\right) = \frac{1}{n^2}\text{Var}(S) = \frac{1}{n^2}n\lambda = \frac{\lambda}{n}
$$

since S is Poisson. Then if n is large, the distribution for $\hat{\lambda}$ is approximately normal. Let $z_{5/2}$ be the value corresponding to 95% confidence, thus the confidence interval is given by

$$
\hat{\lambda} \pm z_{5/2}\sqrt{\frac{\hat{\lambda}}{n}}.
$$

$\square$

(d) Using R, generate a vector of 1000 Poisson random variables with parameter $\lambda = 4$ (you can use the rpois command for this). Now suppose you had this vector of data, but you did not, in fact, know λ. Generate a specific 95% confidence interval for λ based on this data.

Solution. After generating the vector, we had $\bar{X} = 4.022 = \hat{\lambda}$. Then, $z_{5/2} = 1.96$, so the confidence interval is given by

$$
\hat{\lambda} \pm z_{5/2}\sqrt{\frac{\hat{\lambda}}{n}} = 4.022 \pm 0.089.
$$

$\square$

(e) Now suppose that you don't see all of the data. What is the only piece of information from the sample that you need in order to generate the confidence interval in the previous part? Do you really need all n values from the sample?

Answer. We only need to know that the distribution was Poisson, so the variance could be calculated properly.

4. Recall that the PDF for a gamma distribution is of the form

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

where α and λ are unknown positive parameters. Recall that the first two moments μ_1 and μ_2 can be expressed in terms of α and λ and are given by

$$\mu_1 = \frac{\alpha}{\lambda}$$

$$\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$$

In lecture, we showed how this system of equations can be successfully inverted to express α, λ in terms of μ_1, μ_2:

$$\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

$$\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

(a) Suppose $\{X_1, \cdots, X_n\}$ are an iid sample from the gamma distribution with parameters α and λ. In terms of these quantities, what is the exact mean and variance of $\hat{\mu}_1$ and $\hat{\mu}_2$? Also what is the exact covariance between $\hat{\mu}_1$ and $\hat{\mu}_2$?

Solution. We have $\mu_1 = E[X_i] = \alpha/\lambda$, and

$$\mu_2 = E[X_i^2] = \frac{\alpha(\alpha+1)}{\lambda^2} = \mathrm{Var}(X_i) + \left(\frac{\alpha}{\lambda}\right)^2$$

so we can calculate

$$E[\hat{\mu}_1] = E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n E[X_i]$$

$$= \frac{1}{n}n\frac{\alpha}{\lambda} = \boxed{\frac{\alpha}{\lambda}}$$

$$\mathrm{Var}(\hat{\mu}_1) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^n \left[\frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2}\right]$$

$$= \frac{1}{n^2}n\cdot\frac{\alpha}{\lambda^2} = \boxed{\frac{\alpha}{n\lambda^2}}$$

$$E[\hat{\mu}_2] = E\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] = \frac{1}{n}\sum_{i=1}^n E[X_i^2]$$

$$= \frac{1}{n}n\frac{\alpha(\alpha+1)}{\lambda^2} = \boxed{\frac{\alpha(\alpha+1)}{\lambda^2}}$$

$$\mathrm{Var}(\hat{\mu}_2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i^2\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i^2) = \frac{1}{n^2}\sum_{i=1}^n \left(E[X_i^4] - E[X_i^2]\right)$$

Here, we calculate $E[X^4]$ by using integrating:

$$E[X^4] = \int_0^\infty x^4 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\lambda^4} \int_0^\infty \frac{\lambda^{\alpha+4}}{\Gamma(\alpha+4)} x^{(\alpha+4)-1} e^{\lambda x}\, dx$$

$$= \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\lambda^4}$$

since the integrand was the density of the gamma distribution $f(x : \alpha+4, \lambda)$, whose integral is 1. Then since $E[X_i^2] = \frac{\alpha(\alpha+1)}{\lambda^2}$, we have

$$\mathrm{Var}(X_i^2) = \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\lambda^4} - \frac{\alpha^2(\alpha+1)^2}{\lambda^4} = \frac{\alpha(\alpha+1)(4\alpha+6)}{\lambda^4}$$

so using this in the sum, we have

$$\mathrm{Var}(\hat{\mu}_2) = \frac{1}{n^2} \sum_{i=1}^n \frac{\alpha(\alpha+1)(4\alpha+6)}{\lambda^4}$$

$$= \boxed{\frac{\alpha(\alpha+1)(4\alpha+6)}{n\lambda^4}}$$

For the covariance between $\hat{\mu}_1, \hat{\mu}_2$, we have

$$\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_2) = \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{j=1}^n X_j^2\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j^2)$$

Since the X_i are independently sampled, it follows that $\mathrm{Cov}(X_i, X_j^2) = 0$ whenever $i \neq j$. Thus the sum becomes

$$\frac{1}{n^2} \sum_{i=1}^n \mathrm{Cov}(X_i, X_i^2) = \frac{1}{n^2} \sum_{i=1}^n (E[X_i^3] - E[X_i]E[X_i^2])$$

where

$$E[X_i^3] = \int_0^\infty x^3 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\alpha(\alpha+1)(\alpha+2)}{\lambda^3} \int_0^\infty \frac{\lambda^{\alpha+3}}{\Gamma(\alpha+3)} x^{(\alpha+3)-1} e^{-\lambda x}\, dx$$

$$= \frac{\alpha(\alpha+1)(\alpha+2)}{\lambda^3}$$

so the covariance is given by

$$\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_2) = \frac{1}{n^2} \sum_{i=1}^n \left(\frac{\alpha(\alpha+1)(\alpha+2)}{\lambda^3} - \frac{\alpha}{\lambda} \cdot \frac{\alpha(\alpha+1)}{\lambda^2}\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \frac{2\alpha(\alpha+1)}{\lambda^3}$$

$$= \boxed{\frac{2\alpha(\alpha+1)}{n\lambda^3}}$$

$\square$

(b) Using the delta method, what is the approximate mean and variance of $\hat{\alpha}$? Considering these expressions as functions of n, which contributes more toward the MSE of $\hat{\alpha}$: the squared bias or the variance, for large values of n?

Solution. Let

$$\hat{\alpha} = f(\hat{\mu}_1, \hat{\mu}_2) = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2}$$

be a function of $\hat{\mu}_1$ and $\hat{\mu}_2$ and let $\hat{\mu}_1 = x$ and $\hat{\mu}_2 = y$ for convenience. Then consider the second order Taylor expansion about the point (μ_1, μ_2) (assume the function and its derivatives are evaluated there):

$$\alpha \approx f + f_x(x - \mu_1) + f_y(y - \mu_2) + \frac{1}{2}\left[f_{xx}(x - \mu_1)^2 + 2f_{xy}(x - \mu_1)(y - \mu_2) + f_{yy}(y - \mu_2)^2\right]$$

where

$$f_{xx} = \frac{2y(3x^2 + y)}{(y - x^2)^3}$$

$$f_{yy} = \frac{2x^2}{(y - x^2)^3}$$

$$f_{xy} = -\frac{2x(x^2 + y)}{(y - x^2)^3}$$

according to Wolfram.

Note that when we take the expectation of the RHS, the first order derivatives can be ignored since $E[\mu_1 - \hat{\mu}_1] = 0$ and similarly for μ_2. Then $E[(\hat{\mu}_1 - \mu_1)^2] = \text{Var}(\hat{\mu}_1)$ and similarly for $\hat{\mu}_2$, and $E[(\hat{\mu}_1 - \mu_1)(\hat{\mu}_2 - \mu_2)] = \text{Cov}(\hat{\mu}_1, \hat{\mu}_2)$. Then, we have $\mu_1 = \alpha/\lambda$ and $\mu_2 = \alpha(\alpha + 1)/\lambda^2$, so

$$E[\hat{\alpha}] = E[f] + \frac{1}{2}\left(f_{xx}E[(x - \mu_1)^2] + 2f_{xy}E[(x - \mu_1)(y - \mu_2)] + f_{yy}E[(y - \mu_2)^2]\right)$$

$$= \frac{\mu_1^2}{\mu_2 - \mu_1^2} + \frac{1}{2}\left(\frac{2\mu_2(3\mu_1^2 + \mu_2)}{(\mu_2 - \mu_1^2)^3}\text{Var}(\hat{\mu}_1) - \frac{4\mu_1(\mu_1^2 + \mu_2)}{(\mu_2 - \mu_1^2)^3}\text{Cov}(\hat{\mu}_1, \hat{\mu}_2) + \frac{2\mu_1^2}{(\mu_2 - \mu_1^2)^3}\text{Var}(\hat{\mu}_2)\right)$$

$$= \alpha + \frac{3(\alpha + 1)}{n}$$

after substituting and algebra.

I gave up on computing the variance. Oh well.

$\square$

(c) Using the delta method, what is the approximate mean and variance of $\hat{\lambda}$? Considering these expressions as functions of n, which contributes more toward the MSE of $\hat{\lambda}$: the squared bias or the variance, for large values of n?

Solution. Similarly to the previous problem, let

$$\hat{\lambda} = f(\hat{\mu}_1, \hat{\mu}_2) = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1}$$

be a function of $\hat{\mu}_1$ and $\hat{\mu}_2$ and let these be x and y for convenience. The second order Taylor expansion about the point (μ_1, μ_2) is given by

$$\hat{\lambda} \approx f + f_x(x - \mu_1) + f_y(y - \mu_2) + \frac{1}{2}\left[f_{xx}(x - \mu_1)^2 + 2f_{xy}(x - \mu_1)(y - \mu_2) + f_{yy}(y - \mu_2)^2\right]$$

9

where

$$f_{xx} = \frac{2x(x^2 + 3y)}{(y - x^2)^3}$$

$$f_{yy} = \frac{2x}{(y - x^2)^3}$$

$$f_{xy} = -\frac{3x^2 + y}{(y - x^2)^3}$$

and the linear terms go to 0 when the expectation is pulled through, so the expectation is given by

$$E[\hat{\lambda}] = E[f] + \frac{1}{2}\left[\frac{2x(x^2 + 3y)}{(y - x^2)^3}\text{Var}(\hat{\mu}_1) - \frac{2(3x^2 + y)}{(y - x^2)^3}\text{Cov}(\hat{\mu}_1, \hat{\mu}_2) + \frac{2x}{(y - x^2)^3}\text{Var}(\hat{\mu}_2)\right]$$

$$= \lambda + \frac{\lambda}{n}\left(\frac{3\alpha + 4}{\alpha}\right)$$

after substituting and algebra. Gave up on the variance for this question too.

□

(d) Using the delta method, what is the approximate covariance between $\hat{\alpha}$ and $\hat{\lambda}$? (Hint: use a linear approximation to relate $\hat{\alpha}$ to $\hat{\mu}_1$ and $\hat{\mu}_2$ and do the same for $\hat{\lambda}$, then compute the covariance between the approximations)

Solution. Use the first order Taylor expansions for $\hat{\alpha}$ at (μ_1, μ_2) assuming $\hat{\alpha} = f(\hat{\mu}_1, \hat{\mu}_2) = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2}$ is a function, and letting x and y substitute for $\hat{\mu}_1$ and $\hat{\mu}_2$, we have

$$\hat{\alpha} \approx \alpha + f_x(x - \mu_1) + f_y(y - \mu_2)$$

$$= \alpha + \frac{2xy}{(y - x^2)^2}(x - \mu_1) - \frac{x^2}{(y - x^2)^2}(y - \mu_2)$$

$$= \alpha + 2\lambda(\alpha + 1)(\hat{\mu}_1 - \mu_1) - \lambda^2(\hat{\mu}_2 - \mu_2)$$

Similarly, for $\hat{\lambda}$, let $\hat{\lambda} = g(\hat{\mu}_1, \hat{\mu}_2) = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2}$ be a function whose first order Taylor Expansion is

$$\hat{\lambda} \approx \lambda + g_x(x - \mu_1) + g_y(y - \mu_2)$$

$$= \lambda + \frac{x^2 + y}{(y - x^2)^2}(x - \mu_1) - \frac{x}{(y - x^2)^2}(y - \mu_2)$$

$$= \lambda + \frac{\lambda^2(\alpha + \lambda)}{\alpha}(\hat{\mu}_1 - \mu_1) - \frac{\lambda^3}{\alpha}(\hat{\mu}_2 - \mu_2)$$

Then the covariance between the approximations is given by

$$\text{Cov}(\hat{\alpha}, \hat{\lambda}) \approx \text{Cov}\left(\alpha + 2\lambda(\alpha+1)(\hat{\mu}_1 - \mu_1) - \lambda^2(\hat{\mu}_2 - \mu_2), \lambda + \frac{\lambda^2(\alpha+\lambda)}{\alpha}(\hat{\mu}_1 - \mu_1) - \frac{\lambda^3}{\alpha}(\hat{\mu}_2 - \mu_2)\right)$$

$$= \text{Cov}\left(2\lambda(\alpha+1)(\hat{\mu}_1 - \mu_1), \frac{\lambda^2(\alpha+\lambda)}{\alpha}(\hat{\mu}_1 - \mu_1)\right)$$

$$+ \text{Cov}\left(2\lambda(\alpha+1)(\hat{\mu}_1 - \mu_1), -\frac{\lambda^3}{\alpha}(\hat{\mu}_2 - \mu_2)\right)$$

$$+ \text{Cov}\left(-\lambda^2(\hat{\mu}_2 - \mu_2), \frac{\lambda^2(\alpha+\lambda)}{\alpha}(\hat{\mu}_1 - \mu_1)\right)$$

$$+ \text{Cov}\left(-\lambda^2(\hat{\mu}_2 - \mu_2), -\frac{\lambda^3}{\alpha}(\hat{\mu}_2 - \mu_2)\right)$$

$$= \frac{2\lambda^3(\alpha+1)(\alpha+\lambda)}{\alpha}\text{Cov}(\hat{\mu}_1, \hat{\mu}_1) - \frac{2\lambda^4(\alpha+1)}{\alpha}\text{Cov}(\hat{\mu}_1, \hat{\mu}_2)$$

$$- \frac{\lambda^4(\alpha+\lambda)}{\alpha}\text{Cov}(\hat{\mu}_2, \hat{\mu}_1) + \frac{\lambda^5}{\alpha}\text{Cov}(\hat{\mu}_2, \hat{\mu}_2)$$

$$= \frac{2\lambda(\alpha+1)}{n}$$

after algebra.

□

5. In lecture, we described how to use simulation to obtain approximate distributions of our estimators and to determine an estimated standard error. In this exercise, we address this. First, consider fixing the values of α, λ in our gamma distribution at $\alpha = 12$ and $\lambda = 4$, and also fixing n, the sample size, at $n = 200$. In R, complete the following steps:

a) First, do the following for 10, 000 trials. In each trial, sample the gamma distribution using a sample of size 200 (use the rgamma() command) and compute $\hat{\mu}_1$ and $\hat{\mu}_2$. Store the results from each trial as an entry in a vector. You will thus obtain a vector of 10, 000 realizations of values of $\hat{\mu}_1$ and 10, 000 realizations of $\hat{\mu}_2$.

b) Next, for each pair of $(\hat{\mu}_1, \hat{\mu}_2)$ values, compute the corresponding values of $\hat{\alpha}$ and $\hat{\lambda}$, and once again store each result as an entry in a vector. You will thus obtain a vector of 10, 000 realizations of values of $\hat{\alpha}$ and 10, 000 of $\hat{\lambda}$. These vectors of realizations represent independent samples from the sampling distributions of these quantities, and we can use them to understand these distributions.

c) Compute the sample mean and variance of the vector of $\hat{\mu}_1$ values and compare with the theoretical values you computed in the previous problem (Problem 2 (a)). Similarly, compute the sample mean and variance of the vector of $\hat{\mu}_2$ values and compare. Finally, compute the sample covariance between the vectors of $\hat{\mu}_1$ and $\hat{\mu}_2$ values and compare.

Solution. The mean of the $\hat{\mu}_1$ values was 2.99994, which matches closely with $\alpha/\lambda = 12/4 = 3$. The variance was 0.003738, which is very close to

$$\frac{\alpha}{n\lambda^2} = \frac{12}{200 \cdot 4^2} = 0.00375.$$

The mean of the $\hat{\mu}_2$ values is 9.7503, which is close to expression $\alpha(\alpha+1)/\lambda^2 = 12(13)/4^2 = 9.75$. The variance was 0.16397, which is approximately

$$\frac{\alpha(\alpha+1)(4\alpha+6)}{n\lambda^4} = \frac{12(13)(54)}{200(4)^4} = 0.1645$$

The covariance between $\hat{\mu}_1$ and $\hat{\mu}_2$ was 0.02446, which is approximately equal to

$$\frac{2\alpha(\alpha+1)}{n\lambda^3} = \frac{2(12)(13)}{200(4)^3} = 0.02625$$

$\square$

d) Compute the sample mean and variance of the vector of $\hat{\alpha}$ values and compare with the values in Problem 2 (b), and compute the sample mean and variance of the vector of $\hat{\lambda}$ values and compare with the values in Problem 2 (c).

Solution. The sample mean of the $\hat{\alpha}$ values is 12.2105, which is close to the value of

$$\alpha + \frac{3(\alpha+1)}{n} = 12 + \frac{3(13)}{200} = 12.195.$$

The sample variance is 1.6295.
The sample mean of the $\hat{\lambda}$ values is 4.0674, which agrees closely with the value of

$$\lambda + \frac{\lambda}{n}\left(\frac{3\alpha+4}{\alpha}\right) = 4 + \frac{4}{200}\left(\frac{3(12)+4}{12}\right) = 4.067$$

The sample variance is 0.1952.

$\square$

e) Compute the sample covariance between the vectors of $\hat{\alpha}$ and $\hat{\lambda}$ values and compare with the theoretical value in Problem 2 (d).

Solution. The sample covariance is 0.5671, which is approximately equal to

$$\frac{2\lambda(\alpha+1)}{n} = \frac{2(12)(5)}{200} = 0.6.$$

$\square$

6. Recall that we said an estimator $\hat{\theta}_n$, which is based on a sample of size n, is *consistent* for a parameter θ_0 if for any $\varepsilon > 0$,

$$P\left(|\hat{\theta}_n - \theta_0| > \varepsilon\right) \to 0$$

as $n \to \infty$. Let $X_i, 1 \le i \le n$ be an iid collection of random variables with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Show that the sample mean $\bar{X}$ is consistent for μ. (Hint: Use Chebyshev's inequality)

Solution. By Chebyshev's Inequality, we have

$$P\left(|\bar{X} - \mu| > k\frac{\sigma}{\sqrt{n}}\right) \le \frac{1}{k^2}$$

for all $k > 0$, where $\sigma/\sqrt{n}$ is the variance of $\bar{X}$. Then, suppose $k = \frac{\varepsilon\sqrt{n}}{\sigma}$, then the inequality becomes

$$P\left(|\bar{X} - \mu| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$

which will tend to 0 as $n \to \infty$ for any choice of ε as desired. $\square$

Chapter 8: Estimation of Parameters and Fitting of Probability Distributions

4. Suppose that X is a discrete random variable with

$$P(X = 0) = \frac{2}{3}\theta$$

$$P(X = 1) = \frac{1}{3}\theta$$

$$P(X = 2) = \frac{2}{3}(1 - \theta)$$

$$P(X = 3) = \frac{1}{3}(1 - \theta)$$

where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$.

(a) Find the method of moments estimate of θ.

Solution. We have

$$\mu_1 = E[X] = 0 \cdot \frac{2}{3}\theta + 1 \cdot \frac{1}{3}\theta + 2 \cdot \frac{2}{3}(1 - \theta) + 3 \cdot \frac{1}{3}(1 - \theta)$$

$$= \frac{7}{3} - 2\theta$$

so solving for θ gives

$$\theta = \frac{7}{6} - \frac{\mu_1}{2}$$

so the method of moments estimate of θ is given by

$$\hat{\theta} = \frac{7}{6} - \frac{\hat{\mu}_1}{2} = \frac{7}{6} - \frac{\bar{X}}{2} = \frac{7}{6} - \frac{3/2}{2} = \boxed{\frac{5}{12}}.$$

□

(b) Find an approximate standard error for your estimate.

Solution. We have

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{7}{6} - \frac{\bar{X}}{2}\right) = \frac{1}{2^2}\mathrm{Var}(\bar{X})$$

We have

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \left[0^2 \cdot \frac{2}{3}\theta + 1^2 \cdot \frac{1}{3}\theta + 2^2\frac{2}{3}(1 - \theta) + 3^2\frac{1}{3}(1 - \theta)\right] - \left(\frac{7}{3} - 2\theta\right)^2$$

$$= \frac{2}{9} + 4\theta - 4\theta^2$$

so the variance of the sample mean is given by

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\frac{2}{9} + 4\theta - 4\theta^2}{n}.$$

Using the estimated value for $\hat{\theta} = \frac{5}{12}$, we have

$$\mathrm{Var}(\bar{X}) = \frac{\frac{2}{9} - 4 \cdot \frac{5}{12} - 4 \cdot \left(\frac{5}{12}\right)^2}{10} = \frac{43}{360}$$

13

Finally,

$$s_{\hat{\theta}} = \sqrt{\mathrm{Var}(\hat{\theta})} = \sqrt{\frac{1}{4} \cdot \frac{43}{360}} = \boxed{\frac{1}{12}\sqrt{\frac{43}{10}}}$$

$\square$

5. Suppose that X is a discrete random variable with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations of X are made: $x_1 = 1, x_2 = 2, x_3 = 2$.

(a) Find the method of moments estimate of θ.

Solution. We have
$$\mu_1 = E[X] = 1 \cdot \theta + 2 \cdot (1 - \theta) = 2 - \theta$$

so

$$\theta = 2 - \mu_1$$
$$\implies \hat{\theta} = 2 - \hat{\mu}_1 = 2 - \bar{X}$$

where $\bar{X} = (1 + 2 + 2)/3 = 5/3$, so the method of moments estimate of θ is given by

$$\hat{\theta} = 2 - \frac{5}{3} = \boxed{\frac{1}{3}}.$$

$\square$

(b) What is the likelihood function?

Solution. The likelihood function is the joint PMF of X_1, X_2, X_3. That is, the likelihood function is
$$f(\theta) = P(X_1 = 1, X_2 = 2, X_3 = 2) = \theta(1 - \theta)^2.$$

$\square$

(c) What is the maximum likelihood estimate of θ?

Solution. The MLE of θ is
$$\operatorname*{argmax}_{\theta} \theta(1 - \theta)^2$$

Since θ is in the closed interval $[0, 1]$, we take the derivative of the likelihood function wrt θ to get

$$f(\theta) = \theta(-2)(1 - \theta) + (1 - \theta)^2 = 3\theta^2 - 4\theta + 1$$

and the local minima occur where this derivative is 0, which occur when

$$3\theta^2 - 4\theta + 1 = (\theta - 1)(3\theta - 1) = 0$$

so $\theta = 1$ and $\theta = 1/3$. There is also an endpoint at $\theta = 0$ to check.
We have $f(0) = f(1) = 0$ and $f(1/3) = 4/27$. By the second derivative test, we have $f''(1/3) = -2$ so $\boxed{\hat{\theta} = 1/3}$ is a local max, and in fact the global max, so it is the MLE.

$\square$

7. Suppose that X follows a geometric distribution,

$$P(X = k) = p(1 - p)^{k-1}$$

and assume an iid sample of size n.

(a) Find the method of moments estimate of p.

Solution. We have

$$\mu_1 = E[X] = 1/p$$

since X is geometric, so

$$p = 1/\mu_1$$
$$\implies \hat{p} = 1/\hat{\mu}_1 = 1/\bar{X}$$

is the method of moments estimate of p.

$\square$

(b) Find the MLE of p.

Solution. The joint PMF is given by

$$f(p) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$$

since X_i are an iid sample. Since the natural log is an increasing function, maximizing this likelihood function is equivalent to maximizing its logarithm, so let

$$\ell(p) = \log f(p) = \log\left(\prod_{i=1}^{n} p(1-p)^{x_i - 1}\right)$$

$$= \sum_{i=1}^{n} \log\left(p(1-p)^{x_i - 1}\right)$$

$$= \sum_{i=1}^{n} [\log p + (x_i - 1)\log(1-p)]$$

$$= n\log p + \log(1-p)\sum_{i=1}^{n}(x_i - 1)$$

$$= n\log p - n\log(1-p) + \log(1-p)\sum_{i=1}^{n} x_i$$

by properties of logarithms. To maximize this, take the derivative with respect to p and solve for p to make the derivative 0:

$$\ell'(p) = \frac{n}{p} + \frac{n}{1-p} - \frac{1}{1-p}\sum_{i=1}^{n} x_i = 0$$

$$n(1-p) + np = p\sum_{i=1}^{n} x_i$$

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \boxed{1/\bar{X}}$$

is the MLE.

$\square$

(c) Find the asymptotic variance of the MLE.

Solution. The asymptotic variance of $\hat{p}$ is given by $\frac{1}{nI(p)}$, where

$$I(p) = -E\left[\frac{\partial^2}{\partial p^2}\log P(X)\right]$$

$$= -E\left[\frac{\partial^2}{\partial p^2}\log[p(1-p)^{X-1}]\right]$$

$$= E\left[\frac{xp^2 - 2p + 1}{(p-1)^2 p^2}\right]$$

$$= \frac{(1/p)p^2 - 2p + 1}{(p-1)^2 p^2}$$

$$= \frac{1}{p^2(1-p)}$$

so the asymptotic variance is

$$\frac{1}{nI(p)} = \boxed{\frac{p^2(1-p)}{n}}$$

$\square$

16. Consider an iid sample of random variables with density function

$$f(x|\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)$$

(a) Find the method of moments estimate of σ.

Solution. We have

$$\mu_1 = \int_{-\infty}^{\infty} x\left[\frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)\right]dx$$

$$= \frac{1}{2\sigma}\left(\int_{-\infty}^{0} xe^{x/\sigma}\,dx + \int_{0}^{\infty} xe^{-x/\sigma}\,dx\right)$$

$$= 0$$

by symmetry. This won't work, so we use

$$\mu_2 = \int_{-\infty}^{\infty} x^2\left[\frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)\right]dx$$

$$= 2\sigma^2$$

according to Wolfram. Then, since $\mu_1 = 0$, we have $\mu_2 = \text{Var}(X) + \mu_1^2 = \text{Var}(X)$. Thus,

$$\sigma = \sqrt{\frac{\mu_2}{2}}$$

$$\implies \hat{\sigma} = \sqrt{\frac{\hat{\mu}_2}{2}} = \sqrt{\frac{s^2}{2}} = \frac{s}{\sqrt{2}}$$

where s^2 is the sample variance. This is the method of moments estimate.

$\square$

(b) Find the MLE of σ.

Solution. We have the joint density

$$g(\sigma) = f(x_1, x_2, \cdots, x_n | \sigma) = \prod_{i=1}^{n} f(x_i | \sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma} \exp\left(-\frac{|x_i|}{\sigma}\right)$$

Taking the natural log of both sides, we have

$$\ell(\sigma) = \log g(\sigma) = \log \left[\prod_{i=1}^{n} \frac{1}{2\sigma} \exp\left(-\frac{|x_i|}{\sigma}\right) \right]$$

$$= \sum_{i=1}^{n} \log\left[\frac{1}{2\sigma} \exp\left(-\frac{|x_i|}{\sigma}\right) \right]$$

$$= \sum_{i=1}^{n} \left[-\log(2\sigma) - \frac{|x_i|}{\sigma} \right]$$

$$= -n\log(2\sigma) - \sigma^{-1} \sum_{i=1}^{n} |x_i|$$

and taking the derivative with respect to σ and setting the result equal to 0:

$$\ell'(\sigma) = -\frac{n}{\sigma} + \sigma^{-2} \sum_{i=1}^{n} |x_i| = 0$$

$$\sigma = \frac{1}{n} \sum_{i=1}^{n} |x_i| = \hat{\sigma}$$

$\square$

(c) Find the asymptotic variance of the MLE.

Solution. We have

$$I(\sigma) = -E\left[\frac{\partial^2}{\partial \sigma^2} \log f(x|\sigma) \right]$$

$$= -E\left[\frac{\partial^2}{\partial \sigma^2} \log\left(\frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) \right) \right]$$

$$= E\left[\frac{2|x| - \sigma}{\sigma^3} \right]$$

$$= \frac{2E[|x|] - \sigma}{\sigma^3}$$

where

$$E[|x|] = \int_{-\infty}^{\infty} |x| \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx$$

$$= \sigma$$

according to Wolfram, so then

$$I(\sigma) = \frac{2\sigma - \sigma}{\sigma^3} = \frac{1}{\sigma^2}$$

so the asymptotic variance is given by $\boxed{\sigma^2/n.}$

$\square$