Math 318 – homework 5 – due 2023-03-03

Problem 1. Let $X_1, X_2, \dots, X_n$ be independent random variables, each with uniform distribution on $[0, 1]$. Let M be the minimum of these random variables.

(a) Find the cumulative distribution function F_M of the random variable M. (Hint: It might be easier to find $1 - F_M$.)
(b) Find the probability density function of M.
(c) Determine the mean and variance of M.
(d) Let $Y = n \cdot M$. Find the probability density function of Y.
(e) Find the limit as $n \to \infty$ of the pdf of Y.

Solution.

(a) We have
$$P(M > t) = P(X_i > t \text{ for all } i) = P(X_1 > t)^n = (1 - t)^n$$
for $t \in [0, 1]$. Therefore the CDF of M is $F(t) = 1 - (1 - t)^n$.

(b) The PDF of M is $F'(t) = n(1 - t)^{n-1}$ for $t \in [0, 1]$.

(c) By integration:
$$E[M] = \int_0^1 x \cdot n(1 - x)^{n-1}\, dx = \frac{1}{n + 1}.$$
(This is calculated using integration by parts.) The second moment is
$$E[M^2] = \int_0^1 x^2 \cdot n(1 - x)^{n-1}\, dx = \frac{2}{(n + 1)(n + 2)}.$$
E.g. this can be found by integrating by parts twice. Therefore
$$\mathrm{Var}[M] = E[M^2] - E[M]^2 = \frac{n}{(n + 1)^2(n + 2)}.$$

(d) Since $P(Y \le t) = P(M \le t/n) = 1 - (1 - t/n)^n$, the PDF of Y is
$$f_Y(t) = (1 - \tfrac{t}{n})^{n-1}$$
for $t \in [0, n]$.

(e) The limit as $n \to \infty$ is e^{-t}. Asymptotically Y is a standard exponential.

Problem 2. Let X be uniform on $[0, 1]$ and Y be uniform on $[-1, 0]$. Compute $\mathrm{Cov}(X, X^2)$ and $\mathrm{Cov}(Y, Y^2)$. (The first is positive and the second is negative, consistent with the fact that X^2 increases when X whereas the opposite is true for Y.)

Solution. We have $E[X^k] = \int_0^1 x^k\, dx = \frac{1}{k+1}$, and similarly $E[Y^k] = \frac{(-1)^k}{k+1}$. Therefore
$$\mathrm{Cov}(X, X^2) = E[X^3] - E[X]E[X^2] = \tfrac{1}{4} - \tfrac{1}{2}\tfrac{1}{3} = \tfrac{1}{12},$$
and
$$\mathrm{Cov}(Y, Y^2) = E[Y^3] - E[Y]E[Y^2] = \tfrac{-1}{4} - \tfrac{-1}{2}\tfrac{1}{3} = -\tfrac{1}{12}.$$

Problem 3. Let X and Y be independent $N(0, 1)$ random variables. Let $U = X + Y$ and $V = X - Y$.

(a) Find the joint probability density function of U and V.
(b) Show that U and V are independent.
(c) What is the marginal distribution of U?

Solution.

(a) Since X, Y are independent, the joint pdf of X, Y is $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$. Using the change of variable $x = (u+v)/2$ and $y = (u-v)/2$ and remembering the Jacobian which is just $\frac{1}{2}$ gives the pdf of U and V to be

$$g(u,v) = \frac{1}{2} \cdot \frac{1}{2\pi} \exp\left(-(\frac{(u+v)^2}{4} + \frac{(u-v)^2}{4})/2\right) = \frac{1}{4\pi}e^{-(u^2+v^2)/4}.$$

(b) U and V are independent since $g(u,v)$ factors as $\frac{1}{\sqrt{4\pi}}e^{-u^2/4}\frac{1}{\sqrt{4\pi}}e^{-v^2/4}$.

(c) From the above we see that each of U and V is $N(0,2)$. Note that each of these is known to be $N(0,2)$ from the start since that is the distribution of a sum of two standard normal variables.

Problem 4. Let X, Y be the length of time a machine works before breaking and the length of time it takes to fix it, and let $Z = X + Y$. Suppose X, Z have the joint pdf

$$f_{X,Z}(x,z) = \lambda^2 e^{-\lambda z} 1_{0 \le x \le z}.$$

(a) Find the density of X.
(b) Find the density of Z.
(c) Find the joint density of X, Y. (Hint: The joint density of X, Y is $\frac{d}{dx}\frac{d}{dy}\mathbb{P}(X \le x, Y \le y)$.)
(d) Find the density of Y.

Solution.

(a) Density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)dz = \int_x^{\infty} \lambda^2 e^{-\lambda z}dz = \lambda e^{-\lambda x},$$

for $x \ge 0$, so that X is $\text{Exp}(\lambda)$.

(b) Density of Z is

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)dx = \int_0^{z} \lambda^2 e^{-\lambda z}dx = \lambda^2 z e^{-\lambda z},$$

for $z \ge 0$, so that Z is $\text{Gamma}(2, \lambda)$.

(c) Since the Jacobian of $(X, Z) = (X, X+Y)$ is 1, we have by a change of variable $Z = X + Y$,

$$f_{X,Y}(x,y) = \lambda^2 e^{-\lambda(x+y)},$$

for $x, y \ge 0$. Therefore X, Y are independent $\text{Exp}(1)$.

Problem 5. Consider a random variable $X \sim \text{Bin}(100, 0.6)$. We are interested in the event $A = \{X \le 50\}$.

(a) Calculate in python $P(A)$ exactly.
(b) Use the CLT (writing X is a sum of 100 Bernoulli variables) to argue that $Z = \frac{X-a}{b}$ is approximately $N(0,1)$ for some a, b.
(c) Use this to estimate $P(A) = P(Z \le t) = \Phi(t)$ for some t, and calculate this numerically.
(d) The event A is the same as $\{X < 51\}$. Write this as $\{Z < t'\}$ and calculate $\Phi(t')$.
(e) We may get a better approximation by taking the median of these two. Convert $A = \{X \le 50.5\}$ to a statement on Z and the normal approximation for that.

Solution.

(a) This is 0.0271 (see notebook).
(b) Since the binomial is a sum of 100 Bernoulli variables, we can use the CLT to get $Z \approx N(0,1)$ where $Z = \frac{X-np}{\sqrt{np(1-p)}}$. In our case $Z = \frac{X-60}{\sqrt{24}}$.
(c) $P(X \le 50) \approx P(Z \le \frac{50-60}{\sqrt{24}}) = \Phi(\frac{-10}{\sqrt{24}}) = 0.02061$. (See notebook.)
(d) $P(X < 51) \approx P(Z < \frac{51-60}{\sqrt{24}}) = \Phi(\frac{-9}{\sqrt{24}}) = 0.0331$.
(e) $P(X \le 50.5) \approx P(Z \le \frac{50.5-60}{\sqrt{24}}) = \Phi(\frac{-9.5}{\sqrt{24}}) = 0.0262$.

Note: Since X can must be an integer, but cannot be between 50 and 51. We see that using the middle threshold 50.5 when convering to the normal variable gives a better approximation than either endpoint.

Problem 6. Suppose that X, Y are independent discrete random variables taking values in $\mathbb{N}$, with p.m.f.'s p_X and p_Y, respectively. We have seen in class that the random variable $X + Y$ has p.m.f. $p_X * p_Y$, the convolution of p_X and p_Y, i.e.,

$$p_{X+Y}(k) = (p_X * p_Y)(k) = \sum_{j \geq 0}^{k} p_X(j) p_Y(k - j).$$

(a) Let $X_1, X_2, \ldots$ be i.i.d. (independent and identically distributed) random variables that are uniform over the set $\{0, 1, 2\}$, and let $S_n = X_1 + \cdots + X_n$. Calculate (in python) and graph the p.m.f.'s of S_n for $n = 1, 2, 3, 4, 5, 10, 50$. Superimpose each of these with the pdf of a normal random variable with the same mean and variance as S_n.

(b) Let $Y_1, Y_2, \ldots$ be i.i.d. random variables such that

$$\begin{aligned} P(Y_i = -1) = P(Y_i = 0) &= 1/15; \\ P(Y_i = 1) &= 11/15; \\ P(Y_i = 2) = P(Y_i = 4) &= 1/15. \end{aligned}$$

Let $T_n = Y_1 + \cdots + Y_n$. Calculate and graph the p.m.f.'s of T_n for the same n's. Superimpose each of these with the pdf of a normal random variable with the same mean and variance as T_n.

Note: See notebook. Note that X,Y have the same mean and variance, so even though they differ, the sum of n copies of X or of Y look about the same, as the CLT implies.

Extra practice problems

Ross, chapter 2: 53,58,63,71,72,76