Math 425, Section 3, Fall 1997: Introduction to Probability

Final Examination: Solutions

1. Let $\mathbf{Z}$ be a standard unit normal random variable, *i.e.*, a normal random variable with mean 0 and variance 1. Let $\mathbf{Y} = \mathbf{Z}^2$. What is the probability density function of $\mathbf{Y}$? ($\mathbf{Y}$ is called the *chi-squared – or* $\chi^{\mathbf{2}}$ *– random variable, with one degree of freedom.*)

Solution:

Use the usual trick of the derivative of the cumulative distribution function: for $y > 0$,

$$f_{\mathbf{Y}}(y) = \frac{d}{dy}\,P(\mathbf{Y} \le y) = \frac{d}{dy}\,P(\mathbf{Z}^2 \le y) = \frac{d}{dy}\,P(-\sqrt{y} \le \mathbf{Z} \le \sqrt{y}).$$

But $P(-\sqrt{y} \le \mathbf{Z} \le \sqrt{y}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-x^2/2}\,dx$, so

$$f_{\mathbf{Y}}(y) = \frac{d}{dy}\,\frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\,e^{-(\sqrt{y})^2/2}\,\frac{d}{dy}\sqrt{y} - \frac{1}{\sqrt{2\pi}}\,e^{-(-\sqrt{y})^2/2}\,\frac{d}{dy}(-\sqrt{y}),$$

by the fundamental theorem of calculus and the chain rule. Simplifying, we get

$$f_{\mathbf{Y}}(y) = \frac{1}{\sqrt{2\pi y}}\,e^{-y/2}, \text{ for } y > 0,$$

while

$$f_{\mathbf{Y}}(y) = 0, \text{ for } y \le 0.$$

2. Let $\mathbf{X_i}$, $i = 1, 2$, be two exponentially distributed random variables with parameter λ, *i.e.*, $\mathbf{X_i} \sim \mathrm{Exp}(\lambda)$, $i = 1, 2$.

(a) Use *moment generating functions* to show that $\mathbf{Y} = \mathbf{X_1} + \mathbf{X_2}$ has a gamma distribution with parameters $(2, \lambda)$.

Solution:

$M_{\mathbf{Y}}(t) = M_{\mathbf{X_1}+\mathbf{X_2}}(t) = E(e^{t\,(\mathbf{X_1}+\mathbf{X_2})}) = E(e^{t\,\mathbf{X_1}})\,E(e^{t\,\mathbf{X_2}})$, this last by the *independence* of $\mathbf{X_1}$ and $\mathbf{X_2}$. Continuing, a glance at the tables gives us that

$$M_{\mathbf{X_1}+\mathbf{X_2}}\,(t) = (\frac{\lambda}{\lambda - t})^2.$$

Another glance at the tables shows that this is the moment generating function of the gamma distribution with parameters $s = 2$ and $\lambda = \lambda$. Since the moment generating function determines a (reasonable) distribution *uniquely*, we have $\mathbf{X_1} + \mathbf{X_2}$ is gamma-distributed with these parameters.

(b) Give an intuitive argument for (a), using the interpretations of exponential and gamma distributed random variables.

Solution:

An exponential random variable with parameter λ models the waiting time for the first arrival of a Poisson process with rate parameter λ. The sum of two independent such random variables should be the waiting time until the second such arrival, which is represented by a gamma distribution with parameters $s = 2$ and $\lambda = \lambda$.

3. We have two boxes, A and B, with five marbles in each. We know one box contains 3 black marbles and 2 red ones, while the other contains 2 black ones and 3 red ones, but we do not know which is which. We draw a marble from box A, note the color and replace it. Then we draw another marble from box A.

(a) What is the probability of picking a black marble the first draw?

(b) If the first marble is black, what is the probability that A is the box with three black marbles?

(c) What is the probabilty that the second marble drawn is black, if the first one is black?

Solution:

Calculate part (a) by conditioning, letting B_1 be the event that the first marble chosen is black, M the event that box A is the one with 3 black marbles, L the event that box A is the box with 2 black marbles. Then

$$P(B_1) = P(B_1 \,|\, M)\, P(M) + P(B_1 \,|\, L)\, P(L) = \frac{3}{5}\cdot\frac{1}{2} + \frac{2}{5}\cdot\frac{1}{2} = \frac{1}{2}.$$

For the second, we use Bayes's theorem:

$$P(M \,|\, B_1) = \frac{P(B_1 \,|\, M)\cdot P(M)}{P(B_1)} = \frac{3}{5}.$$

Finally, let B_2 denote the event "the second marble drawn is black". We want to calculate the conditional probability

$$P(B_2 \,|\, B_1) = P(B_2 \,|\, M, B_1)\, P(M \,|B_1) + P(B_2 \,|\, L, B_1)\, P(L \,|\, B_1) = \frac{3}{5}\cdot\frac{3}{5} + \frac{2}{5}\cdot\frac{2}{5} = \frac{13}{25},$$

where we use that $P(L \,|\, B_1) = \frac{2}{5}$, which follows from $P(M \,|\, B_1) = \frac{3}{5}$, above.

4. A trained flea sits on the real line at $x = 4$, and his master begins flipping a fair coin. Each time the coin shows a head, the flea hops one unit to the right, each time a tail shows he hops one unit to the left. Let $\mathbf{X}_n$ be the random variable "the flea's position after n flips". So, for example, if the first flip is a head, then $\mathbf{X}_1 = 4 + 1 = 5$.

(a) What is the probability mass function of $\mathbf{X}_3$?

(b) What is the probability that $\mathbf{X}_2 = 3$?

(c) On average, where do you expect the flea to be after four coin flips?

Bonus: What is the probability that the flea will reach the origin $x = 0$ before he reaches $x = 10$? Why?

Solution:

After n flips, the flea's *position* is given by $\mathbf{X_n} = 4 + \mathbf{D_n}$, where $\mathbf{D_n}$ is its *displacement* from $x = 4$ arising from the coin flips. Let $\mathbf{F_i}$ be independent Bernoulli random variables with success probability $\frac{1}{2}$, representing the fair coin flips. The basic equation is

$$\mathbf{D_n} = \sum_{i=1}^{i=n} 2\mathbf{F_i} - 1 = 2\mathbf{B_n} - n,$$

where $\mathbf{B_n}$ is a binomial random variable for n fair flips. In any case, the probability mass function of $\mathbf{X_3}$ is given as follows:

$$\begin{aligned} p_{\mathbf{X_3}}(i) &= \tfrac{1}{8}, \text{ if } i = 1, 7, \\ &= \tfrac{3}{8}, \text{ if } i = 3, 5, \\ &= 0, \text{ all other } i. \end{aligned}$$

As for $P(\mathbf{X_2} = 3)$, by the previous argument, $\mathbf{X_2}$ can only be equal to $2, 4$, and 6, so $P(\mathbf{X_2} = 3) = 0$. Part (c) asks for $E(\mathbf{X_4}) = E(4 + 2\,\mathbf{B_4} - 4) = 2\,E(\mathbf{B_4}) = 4$. In fact, you always expect the flea to be at $x = 4$, after any number of flips, since it is equally likely for him to go left as right at any stage. Finally, for the bonus, notice that the question is the same as the *gambler's ruin problem* as in the textbook, *pp.* 90 - 93 (fifth edition). More exactly, the probability is the same as the probability of ruin for gambler A who flips a fair coin against gambler B, where A's initial stake is 4 units and B's initial stake is 6. Thus the probability the flea will arrive at $x = 0$ before he arrives at $x = 10$ is $\frac{6}{10}$.

5. Fifty numbers are rounded off to the nearest integer and then summed. If the individual round-off errors are uniformly distributed over the interval $(-0.5, 0.5)$, what is the probability that the resultant sum differs from the exact sum (*i.e.,* the sum we would have gotten if we had not rounded off) by more than 3? Be sure to explain the method you use.

Solution:

Let $\mathbf{Y_i}$ be the fifty numbers, and let $\mathbf{X_i}$ be their round-offs. Then $\mathbf{E_i} = \mathbf{Y_i} - \mathbf{X_i}$ are the errors, so $\mathbf{E_i}$ is uniformly distributed over $(-0.5, 0.5)$. The question asks for the probability

$$P(\mid \sum_{i=1}^{i=50} \mathbf{Y_i} - \sum_{i=1}^{i=50} \mathbf{X_i} \mid > 3) = P(\mid \sum_{i=1}^{i=50} \mathbf{E_i} \mid > 3).$$

Now you can use the central limit theorem to approximate by a normal random variable. You need to know the mean and variance of $\mathbf{E_i}$. The mean is 0, and the variance is $\frac{1}{12}$. Hence, $P(\mid \sum_{i=1}^{i=50} \mathbf{E_i} \mid > 3)$ is approximately equal to

$$P(\mid Z \mid > \frac{3-0}{\sqrt{50 \cdot \frac{1}{12}}}) = P(Z > \frac{3-0}{\sqrt{50 \cdot \frac{1}{12}}}) + P(Z < -\frac{3-0}{\sqrt{50 \cdot \frac{1}{12}}}) = 2\,(1 - \Phi(1.470)) = 0.1416,$$

where $\mathbf{Z}$ is a unit normal random variable.

6. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are independent geometric random variables with the same parameter p. What is the conditional probability mass function $P(\mathbf{X} = i \mid \mathbf{X} + \mathbf{Y} = n)$? Give an intuitive argument as well for this result.

Solution:

Given that the sum $\mathbf{X}+\mathbf{Y}=n$, the distribution of $\mathbf{X}$ is uniform over $1, 2, ..., n-1$:

$$\begin{aligned} P(\mathbf{X}=i \,|\, \mathbf{X}+\mathbf{Y}=n) &= \frac{P(\mathbf{X}=i,\, \mathbf{X}+\mathbf{Y}=n)}{P(\mathbf{X}+\mathbf{Y}=n)} \\ &= \frac{P(\mathbf{X}=i,\, \mathbf{Y}=n-i)}{P(\mathbf{X}+\mathbf{Y}=n)} \\ &= \frac{P(\mathbf{X}=i)\, P(\mathbf{Y}=n-i)}{P(\mathbf{X}+\mathbf{Y}=n)} \\ &= \frac{p\,(1-p)^{i-1}\; p\,(1-p)^{n-i-1}}{(n-1)\,p^2\,(1-p)^{n-2}} \\ &= \frac{1}{n-1} \end{aligned}$$

We have used the fact that the random variable $\mathbf{X}+\mathbf{Y}$ has a negative binomial distribution with parameters $2, p$. An intuitive argument could be given that $\mathbf{X}+\mathbf{Y}$ represents the waiting time until a second head is flipped, when repeatedly flipping a coin with bias p. If we know that the second head is obtained on the n-th flip, then there seems no reason to expect the first head to occur at any one of the first $n-1$ flips with greater likelihood than at any other, *i.e.,* we should expect the first head to be uniformly distributed over the $n-1$ possibilities.

7. Consider a line segment of length L, which we will consider to be the interval $[0, L]$ in the real line. Let $\mathbf{X}$ and $\mathbf{Y}$ be random points on this line segment, where $\mathbf{X}$ is uniformly distributed on $[0, L/2]$ and $\mathbf{Y}$ is uniformly distributed on the interval $[L/2, L]$. What is the probability that $\mathbf{X}$ is closer to $\mathbf{Y}$ than to the origin?

Solution:

The joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ is uniform on the square $S = \{(x, y) \,|\, 0 \le x \le L/2,\, L/2 \le y \le L\}$. Our condition is that $\mathbf{Y}-\mathbf{X} < \mathbf{X}$, that is, $\mathbf{Y} < 2\,\mathbf{X}$. The probabil;ity we want to know is thus the portion of the area of S where $L/2 \le y < 2x \le L$. This is easily seen in a picture to be $\frac{1}{4}$.