Worksheet 18 Solution

Mean, Median and Variance of Continuous RVs

1. For each PDF, calculate the mean and the median:

   (a) $f(x) = x^{-2}$ for $\frac{1}{2} \le x \le 1$ and $f(x) = 0$ otherwise.

   $$E[X] = \int_{\frac{1}{2}}^{1} xf(x)dx = \int_{\frac{1}{2}}^{1} \frac{1}{x}dx = \ln x|_{\frac{1}{2}}^{1} = -\ln\frac{1}{2} = \ln 2$$
   $$\int_{\frac{1}{2}}^{m} f(x)dx = \frac{1}{2} \Rightarrow \int_{\frac{1}{2}}^{m} \frac{1}{x^2}dx = \frac{1}{2} \Rightarrow -\left.\frac{1}{x}\right|_{\frac{1}{2}}^{m} = \frac{1}{2} \Rightarrow -\frac{1}{m} + 2 = \frac{1}{2} \Rightarrow m = \frac{2}{3}$$

   (b) $f(x) = x(3x^2 + \frac{1}{2})$ for $0 < x \le 1$ and $f(x) = 0$ otherwise.

   $$E[X] = \int_0^1 xf(x)dx = \int_0^1 x^2(3x^2 + \frac{1}{2})dx = \left.\left(\frac{3x^5}{5} + \frac{x^3}{6}\right)\right|_0^1 = \frac{3}{5} + \frac{1}{6} = \frac{23}{30}$$
   $$\int_0^m f(x)dx = \frac{1}{2} \Rightarrow \int_0^m x(3x^2 + \frac{1}{2})dx = \frac{1}{2} \Rightarrow \left.\left(\frac{3x^4}{4} + \frac{x^2}{4}\right)\right|_0^m = \frac{1}{2}$$
   $$\Rightarrow \frac{3m^4}{4} + \frac{m^2}{4} = \frac{1}{2} \Rightarrow 3m^4 + m^2 - 2 = 0 \Rightarrow m^2 = \frac{2}{3} \Rightarrow m = \sqrt{\frac{2}{3}}$$

   (c) $f(t) = \lambda e^{-\lambda t}$ for $t \ge 0$ and $f(t) = 0$ otherwise.

   $$E[X] = \int_0^\infty tf(t)dt = \int_0^\infty t\lambda e^{-\lambda t}dt = \left(-te^{-\lambda t}\right)\Big|_0^\infty - \int_0^\infty \left(-e^{-\lambda t}\right)dt = \left.\left(-\frac{e^{-\lambda t}}{\lambda}\right)\right|_0^\infty = \frac{1}{\lambda}$$
   $$\int_0^m f(t)dt = \frac{1}{2} \Rightarrow \int_0^m \lambda e^{-\lambda t}dt = \frac{1}{2} \Rightarrow \left(-e^{-\lambda t}\right)\Big|_0^m = \frac{1}{2} \Rightarrow 1 - e^{-\lambda m} = \frac{1}{2} \Rightarrow m = \frac{\ln 2}{\lambda}$$

   (d) $f(x) = c(1 - x^2)$ for $-1 < x < 1$ and $f(x) = 0$ otherwise. Does the answer depend on c? Why?
   We can do integration to find out that $E[X] = m = 0$. We observe that $f(x)$ is symmetric about 0 and the domain of x is also symmetric about 0, so we can conclude that the mean and the median must be at the symmetric point.

2. **Chromosomal recombination** is a process by which two chromosomes join together and exchange DNA. The point along the DNA at which the join occurs is randomly located. Suppose X is a RV denoting the location with $0 \le X \le 2$. In an experiment, $E[X] = 1$ and $Var[X] = \frac{1}{3}$. Are the findings consistent with the hypothesis that all locations along the chromosome are equally likely to contain the join point? Explain.

   Hypothesis: $X \sim Uniform(0, 2) \Rightarrow E[X] = \frac{0+2}{2} = 1$. $Var[X] = \frac{(2-0)^2}{12} = \frac{1}{3}$. Thus, the findings are consistent with the hypothesis.

Midterm 2 Review

Worksheet 18 Solution

1. Find a, b or c given the PDF, then find the CDF.

   (a) $f(x) = c(1 - x^2)$ for $-1 < x < 1$ and $f(x) = 0$ otherwise.

   $$\int_{-1}^{1} c(1-x^2)dx = 1 \Rightarrow \left(cx - \frac{cx^3}{3}\right)\Big|_{-1}^{1} = 1 \Rightarrow c = \frac{3}{4}$$

   (b) $f(x) = c/x^2$ for $x > 10$ and $f(x) = 0$ otherwise.

   $$\int_{10}^{\infty} \frac{c}{x^2}dx = 1 \Rightarrow \left(-\frac{c}{x}\right)\Big|_{10}^{\infty} = 1 \Rightarrow c = 10$$

   (c) $f(x) = a + bx^2$ for $0 \leq x \leq 1$ and $f(x) = 0$ otherwise, given $E(X) = \frac{3}{5}$.

   $$\frac{3}{5} = E[X] = \int_0^1 xf(x)dx = \int_0^1 x(a+bx^2)dx = \left(\frac{ax^2}{2} + \frac{bx^4}{4}\right)\Big|_0^1 = \frac{a}{2} + \frac{b}{4}$$
   $$1 = \int_0^1 f(x)dx = \left(ax + \frac{bx^3}{3}\right)\Big|_0^1 = a + \frac{b}{3}$$
   $$\Rightarrow \quad a = \frac{3}{5}, \quad b = \frac{6}{5}$$

2. Suppose you take a random sample of 10 tickets without replacement from a box containing 20 red tickets and 30 blue tickets.

   (a) What is the chance of getting exactly 4 red tickets?

   $X \sim Hypergeometric(50, 20, 10)$
   $P(X = 4) = \dfrac{C(20,4) * C(30,6)}{C(50,10)}$

   (b) Repeat (a) for sampling with replacement.

   $X \sim Binomial(10, 0.4)$
   $P(X = 4) = C(10,4) * 0.4^4 * 0.6^6$

3. Suppose that we observed 10 frogs in a pond during the obsevation period of 100 days. Find the Poisson approximation to the probability of observing $X = k$ frogs each day. Using that approximation to calculate the probability that

   (a) you observe precisely one frog today?

   On average, each day we observe $\lambda = 10/100 = 0.1$ frogs.
   Approximating $X \sim Poisson(\lambda)$ gives $P(X = k) = 0.1^k * \dfrac{e^{-0.1}}{k!}$
   Thus, $P(X = 1) = 0.1^1 * \dfrac{e^{-0.1}}{1!} = 0.1 * e^{-0.1}$

(b) you observe more than one frog today?
$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-0.1} - 0.1e^{-0.1}$

(c) you observe no frogs today?
$P(X = 0) = 0.1^0 * \dfrac{e^{-0.1}}{0!} = e^{-0.1}$

4. Suppose that a test for opium use has a 2% false positive rate and a 5% false negative rate. That is, 2% of people who do not use opium test positive for opium, and 5% of opium users test negative for opium. Furthermore, suppose that 1% of people actually use opium.

Let A: use opium, B: test positive. Given:
$P(B|\bar{A}) = 0.02, \quad P(\bar{B}|A) = 0.05$
$P(A) = 0.01$

(a) Find the probability that someone who tests negative for opium use does not use opium.

$$\begin{aligned} P(\bar{A}|\bar{B}) &= \frac{P(\bar{B}|\bar{A}) * P(\bar{A})}{P(\bar{B}|\bar{A}) * P(\bar{A}) + P(\bar{B}|A) * P(A)} = \frac{1}{1 + \frac{P(\bar{B}|A)*P(A)}{P(\bar{B}|\bar{A})*P(\bar{A})}} \\ &= \frac{(1-0.02)(1-0.01)}{(1-0.02)(1-0.01) + 0.05 * 0.01} = 0.9995 \end{aligned}$$

(b) Find the probability that someone who tests positive for opium use actually uses opium.

$$\begin{aligned} P(A|B) &= \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})} = \frac{1}{1 + \frac{P(B|\bar{A})*P(\bar{A})}{P(B|A)*P(A)}} \\ &= \frac{(1-0.05)0.01}{(1-0.05)0.01 + 0.02(1-0.01)} = 0.3242 \end{aligned}$$

Source: some from Stewart's *Biocalculus*, the others from internet.