STA2001 Assignment 5

1. (3.1-3). Customers arrive randomly at a bank teller's window. Given that one customer arrived during a particular 10-minute period, let X equal the time within the 10 minutes that the customer arrived. If X is $U(0, 10)$, find
(a) The pdf of X.
(b) $P(X \geq 8)$.
(c) $P(2 \leq X < 8)$.
(d) $E(X)$.
(e) $Var(X)$.

Solution:

(a) The pdf is $f(x) = \frac{1}{10}, \quad 0 < x < 10$
(b) $P(X \geq 8) = \int_8^{10} \frac{1}{10} dx = 0.2$
(c) $P(2 \leq X < 8) = \int_2^8 \frac{1}{10} dx = 0.6$
(d) $E(X) = \frac{0+10}{2} = 5$
(e) $\text{Var}(X) = \frac{(10-0)^2}{12} = \frac{25}{3}$

2. (3.1-4). If the mgf of X is

$$M(t) = \frac{e^{5t} - e^{4t}}{t}, t \neq 0 \text{ and } M(0) = 1$$

find (a) $E(X)$, (b) $Var(X)$, and (c) $P(4.2 < X \leq 4.7)$.

Solution:

By matching the moment generating function, we know that $X \sim U(4, 5)$. Hence, we have
(a) $E(X) = \frac{1}{2}(4 + 5) = \frac{9}{2}$
(b) $\text{Var}(X) = \frac{(5-4)^2}{12} = \frac{1}{12}$
(c) $P(4.2 < X \leq 4.7) = \int_{4.2}^{4.7} 1dx = 0.5$
Note that it's important to remember the mgfs of distributions you have learned so that you could identify the distribution by recognising the corresponding mgf.

3. (3.1-5). Let Y have a uniform distribution $U(0, 1)$, and let

$$W = a + (b - a)Y, \quad a < b.$$

(a) Find the cdf of W.
Hint: Find $P[a + (b - a)Y \leq w]$.
(b) How is W distributed?

Solution: (a) Note that pdf of Y is just 1. Let G be the cdf of W, then for $a < w < b$

$$G(w) = P(W \leq w) = P(a+(b-a)Y \leq w) = P\left(Y \leq \frac{w-a}{b-a}\right) = \int_0^{\frac{w-a}{b-a}} 1dy = \frac{w-a}{b-a}$$

and $G(w) = 0$ for $w \leq a$ and $G(w) = 1$ for $w \geq b$.

(b) From the cdf of W in (a), we know that $W \sim U(a, b)$

4. (3.1-6). A grocery store has n watermelons to sell and makes \$1.00 on each sale. Say the number of consumers of these watermelons is a random variable with a distribution that can be approximated by

$$f(x) = \frac{1}{200}, \quad 0 < x < 200,$$

a pdf of the continuous type. If the grocer does not have enough watermelons to sell to all consumers, she figures that she loses \$5.00 in goodwill from each unhappy customer. But if she has surplus watermelons, she loses 50 cents on each extra watermelon. What should n be to maximize profit? Hint: If $X \leq n$, then her profit is $(1.00)X + (-0.50)(n - X)$; but if $X > n$, her profit is $(1.00)n + (-5.00)(X - n)$. Find the expected value of profit as a function of n, and then select n to maximize that mfunction.

Solution:

From the hint, we let $g(X)$ be the profit function and write

$$g(X) = \begin{cases} (1.00)X + (-0.50)(n - X), & 0 \leq X \leq n \\ (1.00)n + (-5.00)(X - n), & 200 \geq X > n \end{cases}$$

where we have used the fact that the continuous random variable X only has positive density on $(0, 200)$

So the mean of $g(X)$ can be further written as

$$\begin{aligned} E(g(X)) &= \int_0^{200} g(x)f(x)dx \\ &= \left(\int_0^n (1.5x - 0.5n)dx + \int_n^{200} (-5x + 6n)dx\right) \cdot \frac{1}{200} \\ &= \frac{1}{200}\left(-\frac{13}{4}n^2 + 1200n - 100000\right) \end{aligned}$$

Using the first order condition (i.e. set the first derivative as zero),

$$\frac{d}{dn}E(g(X)) = \frac{d}{dn}\frac{1}{200}\left(-\frac{13}{4}n^2 + 1200n - 100000\right) = \frac{1}{200}\left(-\frac{13}{2}n + 1200\right) = 0$$

and solve the equation we yields

$$n = \frac{2400}{13} = 184.6153 \approx 185$$

Note that we could check that it's indeed a maximum point as the function $E(g(X))$ is concave in n, i.e. its second derivative is negative.

5. (3.1-8). For each of the following functions, (i) find the constant c so that $f(x)$ is a pdf of a random variable X, (ii) find the cdf, $F(x) = P(X \leq x)$, (iii) sketch graphs of the pdf $f(x)$ and the distribution function $F(x)$, and (iv) find μ and σ^2:
(a) $f(x) = (3/16)x^2, -c < x < c$.
(b) $f(x) = c/\sqrt{x}, 0 < x < 1$. Is this pdf bounded?

Solution: (a)

(i) The pdf must satisfy the equation

$$\int_{-c}^{c}(3/16)x^2dx = \frac{3}{48}x^3\bigg|_{-c}^{c} = \frac{3}{48}c^3 + \frac{3}{48}c^3 = 1$$

Solving the equation we have $c = 2$.

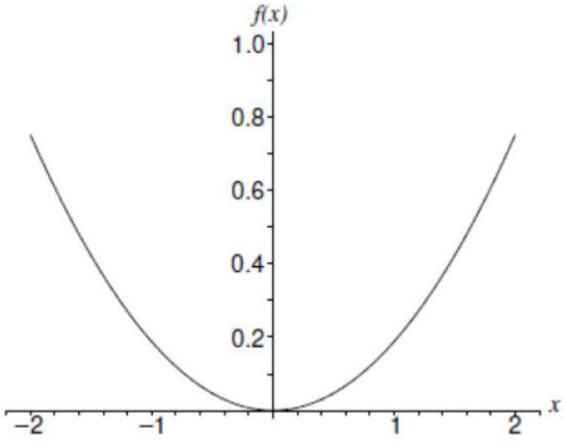
(ii) We compute the integral,

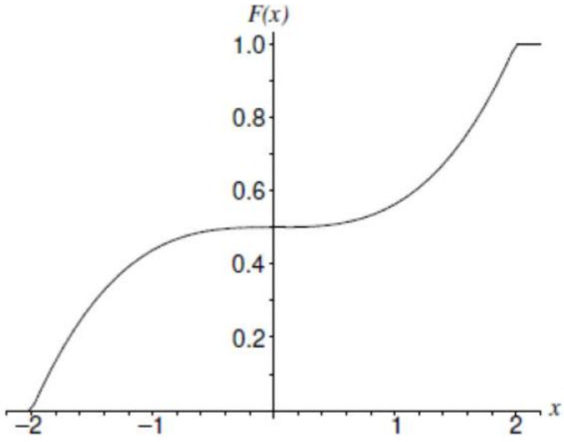
$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-2}^{x}(3/16)t^2dt = \frac{t^3}{16}\bigg|_{-2}^{x} = \frac{x^3}{16} + \frac{1}{2}$$

Summarizes it gives us

$$F(x) = \begin{cases} 0, & -\infty < x < -2 \\ \frac{x^3}{16} + \frac{1}{2}, & -2 \leq x < 2 \\ 1, & 2 \leq x < \infty \end{cases}$$

(iii) The graphs are given below:





(iv) We compute the mean and variance according to the definition,

$$\mu = E(X) = \int_{-2}^{2} x \frac{3}{16} x^2 dx = 0$$
$$\sigma^2 = E\left(X^2\right) - (E(X))^2 = E\left(X^2\right) = \int_{-2}^{2} x^2 \frac{3}{16} x^2 dx = \frac{12}{5}$$

(b)

(i) The pdf must satisfy the equation

$$\int_0^1 \frac{c}{\sqrt{x}} dx = 2c\sqrt{x}\Big|_0^1 = 2c = 1$$

Solving the equation we have $c = 1/2$.

The pdf $\frac{1}{2\sqrt{x}}$ for $0 < x < 1$, is unbounded, as for any given real number M we could always choose a sufficient small $\epsilon > 0$ such that $|f(\epsilon)| > M$.

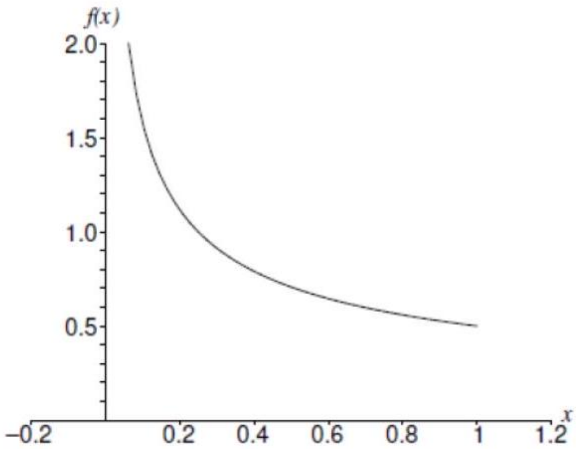
(ii) We compute the integral

$$F(x) = \int_{-\infty}^{x} f(t) dt = \int_0^x \frac{1}{2\sqrt{t}} dt = \sqrt{t}\Big|_0^x = \sqrt{x}$$

Summarizes it gives us

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ \sqrt{x}, & 0 \le x < 1 \\ 1, & 1 \le x < \infty \end{cases}$$

(iii) The graphs are given below:



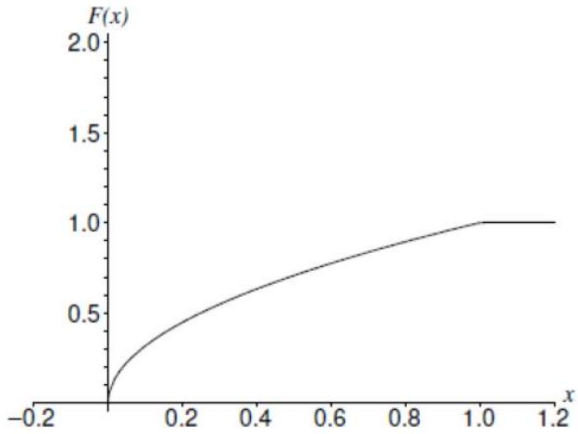


Figure 3.1–8: **(c)** Continuous distribution pdf and cdf

(iv) We compute the mean and variance according to the definition,

$$\mu = E(X) = \int_0^1 x \frac{1}{2\sqrt{x}} dx = \frac{1}{3}$$
$$E\left(X^2\right) = \int_0^1 x^2 \frac{1}{2\sqrt{x}} dx = \frac{1}{5}$$
$$\sigma^2 = E\left(X^2\right) - (E(X))^2 = \frac{1}{5} - \left(\frac{1}{3}\right)^2 = \frac{4}{45}$$

6. (3.1-16). Let $f(x) = (x+1)/2, -1 < x < 1$. Find (a) $\pi_{0.64}$, (b) $q_1 = \pi_{0.25}$, and (c) $\pi_{0.81}$.

Solution: (a) First, we compute the cdf

$$F(x) = \int_{-1}^{x} \frac{t+1}{2} dt = \frac{1}{2}\left(\frac{1}{2}(t+1)^2\right)\Big|_{-1}^{x} = \frac{(x+1)^2}{4}, \quad -1 < x < 1$$

The 64 th percentile $\pi_{0.64}$ satisfies the equation

$$F\left(\pi_{0.64}\right) = \frac{\left(\pi_{0.64}+1\right)^2}{4} = 0.64$$

Solve this equation yields,

$$\pi_{0.64} = \sqrt{4 \times 0.64} - 1 = 0.6$$

(b) Similarly, solve the equation

$$F\left(\pi_{0.25}\right) = \frac{\left(\pi_{0.25}+1\right)^2}{4} = 0.25$$

gives us $\pi_{0.25} = 0$.

(c) Similarly, solve the equation

$$F(\pi_{0.81}) = \frac{(\pi_{0.81}+1)^2}{4} = 0.81$$

gives us $\pi_{0.81} = 0.8$.

7. (3.1-21). Let $X_1, X_2, \cdots, X_k$ be random variables of the continuous type, and let $f_1(x), f_2(x), \cdots, f_k(x)$ be their corresponding pdfs, each with sample space $S = (-\infty, \infty)$. Also, let $c_1, c_2, \cdots, c_k$ be nonnegative constants such that $\sum_{i=1}^{k} c_i = 1$.
(a) Show that $\sum_{i=1}^{k} c_i f_i(x)$ is a pdf of a continuous-type random variable on S.
(b) If X is a continuous-type random variable with pdf $\sum_{i=1}^{k} c_i f_i(x)$ on S, $E(X_i) = \mu_i$, and $Var(X_i) = \sigma_i^2$ for $i = 1, \cdots, k$, find the mean and the variance of X.

Solution: (a) We check the conditions that a pdf of a continuous-type random variable must satisfy.

Firstly, since $c_i \geq 0$ and $f_i(x) \geq 0$ for any $x \in S$, we have

$$\sum_{i=1}^{k} c_i f_i(x) \geq 0$$

Secondly, integrating over S we have

$$\int_{-\infty}^{\infty} \sum_{i=1}^{k} c_i f_i(x) dx = \sum_{i=1}^{k} c_i \int_{-\infty}^{\infty} f_i(x) dx = \sum_{i=1}^{k} c_i \cdot 1 = 1$$

where we have used fact that $f_i(x)$ are pdf and $\sum_{i=1}^{k} c_i = 1$.

These two conditions are sufficient enough for a function to be a legitimate pdf, and thus there must be a continuous type random variable has the pdf $\sum_{i=1}^{k} c_i f_i(x)$.

(b) Since for $i = 1, 2, \ldots, k$ we have

$$\int_{-\infty}^{\infty} x f_i(x) dx = \mu_i$$
$$\int_{-\infty}^{\infty} (x-\mu_i)^2 f_i(x) dx = \sigma_i^2$$
$$\int_{-\infty}^{\infty} x^2 f_i(x) dx = \mu_i^2 + \sigma_i^2$$

we can calculate the mean and the variance of X in the following way,

$$E(X)=\int_{-\infty}^{\infty} x \sum_{i=1}^{k} c_i f_i(x) d x=\sum_{i=1}^{k} c_i \int_{-\infty}^{\infty} x f_i(x) d x=\sum_{i=1}^{k} c_i \mu_i$$
$$E\left(X^2\right)=\int_{-\infty}^{\infty} x^2 \sum_{i=1}^{k} c_i f_i(x) d x=\sum_{i=1}^{k} c_i \int_{-\infty}^{\infty} x^2 f_i(x) d x=\sum_{i=1}^{k} c_i\left(\mu_i^2+\sigma_i^2\right)$$
$$\operatorname{Var}(X)=E\left(X^2\right)-[E(X)]^2=\sum_{i=1}^{k} c_i\left(\mu_i^2+\sigma_i^2\right)-\left(\sum_{i=1}^{k} c_i \mu_i\right)^2$$

8. (3.2-1). What are the pdf, the mean, and the variance of X if the moment-generating function of X is given by the following?
(a) $M(t)=\frac{1}{1-3t}, t<1/3$.
(b) $M(t)=\frac{3}{3-t}, t<3$.
Solution: (a) By matching the moment generating function, we know that X has an exponential distribution with parameter $\theta=3$. The corresponding pdf is given by

$$f(x)=\frac{1}{3} e^{-x/3}, \quad 0 \leq x<\infty$$

and zero elsewhere.

Since it's an exponential distribution, its mean and variance are

$$E(\mathrm{X})=\theta=3, \quad \operatorname{Var}(X)=\theta^2=9$$

(b) By matching the moment generating function, we know that X has an exponential distribution with parameter $\theta=1/3$. The corresponding pdf is given by

$$f(x)=3 e^{-3x}, \quad 0 \leq x<\infty$$

and zero elsewhere.

Since it's an exponential distribution, its mean and variance are

$$E(X)=\theta=\frac{1}{3}, \quad \operatorname{Var}(X)=\theta^2=\frac{1}{9}$$

Note: we could also compute the moments (and thus the variance) by using the property $E\left(X^k\right)=\left.M^{(k)}(t)\right|_{t=0}$, for positive integers k and a given mgf. This is a general method and it's important in our course.

9. (3.2-3). Let X have an exponential distribution with mean $\theta>0$. Show that

$$P(X>x+y \mid X>x)=P(X>y).$$

Solution: For $x > 0$ and $y < 0$, it's obvious that $P(X > x + y \mid X > x) = 1 = P(X > y)$, as the cdf of an exponential distribution says that $P(X > y) = 1$ for any $y < 0$.

Now we consider $x > 0$ and $y > 0$. By Bayes' theorem,

$$P(X > x + y \mid X > x) = \frac{P(X > x + y)}{P(X > x)} = \frac{e^{-(x+y)/\theta}}{e^{-x\theta}} = e^{-y\theta} = P(X > y)$$

which completes the proof.

10. A random variable X with parameter $\theta \in \mathbb{R}$ belongs to the exponential family if its probability mass function or probability density function can be written as

$$f(x) = h(x)e^{\theta x - A(\theta)},$$

where $h(x)$ and $A(\theta)$ are some known functions. Note that $h(x)$ does not depend on θ and $A(\theta)$ does not depend on x.

(a1) Show that a Poisson distribution with mean $\lambda > 0$ belongs to the exponential family by identifying appropriate $h(x)$, θ and $A(\theta)$.

(a2) Assume that X is a continuous random variable that belongs to the exponential family. Show that the moment generating function $M(t)$ of X can be written as

$$M(t) = E(e^{tX}) = e^{A(\theta+t)-A(\theta)}.$$

(a3) Continuing part (a2), and assume that $A(\theta)$ is twice differentiable so that its first and second derivative with respect to θ exist. Show that the mean and variance of X are

$$E(X) = \frac{d}{d\theta}A(\theta), \quad \mathrm{Var}(X) = \frac{d^2}{d\theta^2}A(\theta).$$

Solution:

Part (a1). The probability mass function of Poisson random variable with mean λ is given by

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{1}{x!}e^{(\ln\lambda)x}e^{-\lambda} = h(x)e^{\theta x - A(\theta)},$$

where we take

$$\begin{aligned} h(x) &= \frac{1}{x!} \\ \theta &= \ln\lambda \\ A(\theta) &= \lambda. \end{aligned}$$

Part (a2).

$$
\begin{aligned}
E(e^{tX}) &= \int e^{tx} h(x) e^{\theta x - A(\theta)} \, dx \\
&= e^{-A(\theta)} \int h(x) e^{(\theta+t)x} \, dx \\
&= e^{A(\theta+t)-A(\theta)} \int h(x) e^{(\theta+t)x - A(\theta+t)} \, dx \\
&= e^{A(\theta+t)-A(\theta)},
\end{aligned}
$$

where in the last equality we use $\int h(x) e^{(\theta+t)x - A(\theta+t)} \, dx = 1$.

Part (a3).

$$
\begin{aligned}
\frac{d}{dt} E(e^{tX}) &= e^{A(\theta+t)-A(\theta)} \frac{d}{dt} A(\theta+t) \\
\frac{d^2}{dt^2} E(e^{tX}) &= e^{A(\theta+t)-A(\theta)} \frac{d^2}{dt^2} A(\theta+t) + e^{A(\theta+t)-A(\theta)} \left(\frac{d}{dt} A(\theta+t) \right)^2.
\end{aligned}
$$

As a result, by taking $t = 0$ above, we have

$$
\begin{aligned}
E(X) &= \frac{d}{dt} E(e^{tX}) \bigg|_{t=0} = \frac{d}{d\theta} A(\theta), \\
E(X^2) &= \frac{d^2}{dt^2} E(e^{tX}) \bigg|_{t=0} = \frac{d^2}{d\theta^2} A(\theta) + \left(\frac{d}{d\theta} A(\theta) \right)^2, \\
\mathrm{Var}(X) &= E(X^2) - E(X)^2 = \frac{d^2}{d\theta^2} A(\theta).
\end{aligned}
$$