MAT2013: Probability and Statistics **Spring 2017**

Problem Set 2

Name: Jisung Lim (2014147040)

Solution

1. Let X be a (absolutely) continuous random variable with probability density f_X. What is the density of $Y = X^2$?
Solution:

Random variable Y is defined by a function of X, which is given by

$$Y = \phi(X) \quad \text{where} \quad \phi(X) = X^2 \tag{2.1}$$

First, to find the density f_Y, we first need to find the distribution function F_Y, of which the density f_Y is defined as a integrand:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\ \mathrm{d}t \tag{2.2}$$

Now, let's consider the probability space $(\mathbb{R}, \mathcal{B}, P_Y)$ which is induced by the random variable Y from 'primitive' probability space, says $(\Omega, \mathcal{A}, \mathbb{P})$. Then, the probability distribution P_Y for any $B \in \mathcal{B}$ is defined by

$$P_Y[Y \in B] = \mathbb{P}[\{\omega \in \Omega : \omega \in Y^{-1}(B)\}] \tag{2.3}$$

and the distribution function $F_Y(y)$ is defined by the probability distribution $P_Y[\cdot]$, given by

$$F_Y(y) = P_Y[Y \in (-\infty, y)] \tag{2.4}$$

Hence, by (2.3) and (2.4), the probability distribution can be expressed as follows

$$\begin{aligned} F_Y(y) &= P_Y[Y \in (-\infty, y)] \\ &= \mathbb{P}[\{\omega \in \Omega : Y(\omega) \in (-\infty, y)\}] \\ &= \mathbb{P}[\{\omega \in \Omega : Y(\omega) < y\}] \end{aligned} \tag{2.5}$$

and then, we can use given knoweledge (2.1), and the definition of the distribution function of X, says $F_X(x)$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}[\{\omega \in \Omega : Y(\omega) < y\}] \\ &= \mathbb{P}[\{\omega \in \Omega : X^2(\omega) < y\}] \\ &= \mathbb{P}[\{\omega \in \Omega : -y^{1/2} < X(\omega) < y^{1/2}\}] \\ &= \mathbb{P}[\{\omega \in \Omega : X(\omega) < y^{1/2}\}] - \mathbb{P}[\{\omega \in \Omega : X(\omega) < -y^{1/2}\}] \\ &= F_X(y^{1/2}) - F_X(y^{-1/2}) \end{aligned} \tag{2.6}$$

where

$$F_X(x) = P_X[X \in (-\infty, x)] \tag{2.7}$$

Since the probability distribution function $f_X(x)$ is given by

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t, \tag{2.8}$$

$F_Y(y)$ can be expressed by

$$\begin{aligned} F_Y(y) &= F_X(y^{1/2}) - F_X(y^{-1/2}) \\ &= \int_{-y^{1/2}}^{y^{1/2}} f_X(t)\,\mathrm{d}t \end{aligned} \tag{2.9}$$

To find probability distribution function f_Y, find the derivative of the distribution function, then we get

$$\begin{aligned} f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} &= \frac{\mathrm{d}}{\mathrm{d}y}\left[F_X(y^{1/2}) - F_X(y^{-1/2})\right] \\ &= f_X(y^{1/2})\frac{\mathrm{d}y^{1/2}}{\mathrm{d}y} - f_X(y^{-1/2})\frac{\mathrm{d}y^{-1/2}}{\mathrm{d}y} \\ &= \frac{1}{2\sqrt{y}}f_X(y^{1/2}) + \frac{1}{2y\sqrt{y}}f_X(y^{-1/2}) \end{aligned} \tag{2.10}$$

Hence,

$$f_Y(y) = \frac{1}{2\sqrt{y}}f_X(y^{1/2}) + \frac{1}{2y\sqrt{y}}f_X(y^{-1/2}) \tag{2.11}$$

■

2. A certain river floods every day. Suppose that the low-water mark is set at 1 and the high-water mark Y has distribution function
$$F_Y(y) = 1 - \frac{1}{y^2}, \quad 1 \le y < \infty. \tag{2.12}$$

(a) Verify that $F_Y(y)$ is a distribution function.
Solution:
First, investigate the convergence of the function.
$$\lim_{y\to\infty} F_Y(y) = \lim_{y\to\infty}\left(1 - \frac{1}{y^2}\right) = 1 - \lim_{y\to\infty}\frac{1}{y^2} = 1 - 0 = 1 \tag{2.13}$$
and since $F_Y(y) = 0 \quad \forall y < 1$, then
$$\lim_{y\to-\infty} F_Y(y) = 0 \tag{2.14}$$
Second, the function must be monotonously none-decreasing, i.e.,
$$\forall a, b \in \mathbb{R}, a < b \Rightarrow F_Y(a) \le F_Y(b) \tag{2.15}$$
and there are three cases.

i. $a < b < 1$; Since the function $F_Y(y) = 0$ $\forall y < 1$, the function satisfies $a < b \Rightarrow F_Y(a) = F_Y(b) = 0$, i.e., monotone none-decreasing in $(-\infty, y)$.

ii. $1 \le a < b$; Since the function $1 - 1/y^2$ is strictly increasing function, it satisfies that $a < b \Rightarrow F_Y(a) < F_Y(b)$, which also satisfies monotone none-decreasing condition.

iii. $a < 1 \le b$; At first, $F_Y(a) = 0$. If we investigate the limit point 1^+, then
$$\lim_{y\to1^+} F_Y(y) = \lim_{y\to1^+}\left(1 - \frac{1}{y^2}\right) = 1 - \lim_{y\to1^+}\frac{1}{y^2} = 1 - 1 = 0, \tag{2.16}$$
Since the function $F_Y(y)$ is strictly increasing function in $[1, \infty)$, $F_Y(b) > 0$ $\forall b \ge 1$. Hence, the function satisfies that $a < b \Rightarrow F_Y(a) < F_Y(b)$. i.e., $F_Y(y)$ is monotone none-decreasing function.

Since the function $F_Y(y)$ is monotone none-decreasing function and satisfies $\lim_{y\to\infty} F_Y(y) = 1$ and $\lim_{y\to-\infty} F_Y(y) = 0$, then the function $F_Y(y)$ is distribution function. ■

(b) Find $f_Y(y)$, the pdf of Y.
Solution:
By definition of probability density function $f_Y(y)$ and distribution function $F_y(y)$
$$F_Y(y) = P_Y(Y \le y) = \int_{-\infty}^{y} f_Y(t)\mathrm{d}t = \int_{1}^{y} f_Y(t)\mathrm{d}t \tag{2.17}$$
By FTC1
$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{2}{y^3} \tag{2.18}$$
■

(c) If the low-water mark is reset at 0 and we use a unit of measurement that is $\frac{1}{10}$ of that given previously, the high-water mark become $Z = 10(Y-1)$. Find $F_Z(z)$.
$$\begin{aligned} F_Z(z) &= P_Z(Z \le z) \\ &= \mathbb{P}(\{\omega \in \Omega : Z \le z\}) \\ &= \mathbb{P}(\{\omega \in \Omega : 10(Y-1) \le z\}) \end{aligned} \tag{2.19}$$

Since $\phi(x) = x/10 + 1$ is monotone increasing function, then $a < b \Rightarrow \phi(a) \leq \phi(b)$.

$$\begin{aligned} F_Z(z) &= \mathbb{P}(\{\omega \in \Omega : 10(Y-1) \leq z\}) \\ &= \mathbb{P}(\{\omega \in \Omega : Y \leq z/10 + 1\}) \\ &= P_Z(Y \leq z/10 + 1) = F_Y(z/10 + 1) \\ &= \begin{cases} 1 - \frac{1}{(z/10+1)^2}, & 1 \leq y < \infty \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{2.20}$$

Since $y \geq 1$ and $z = 10(y-1)$, then $z \geq 0$ and therefore

$$F_Z(z) = \begin{cases} 1 - \frac{100}{(z+10)^2}, & 0 \leq z < \infty \\ 0, & \text{otherwise} \end{cases} \tag{2.21}$$

3. (Geometric Distribution) Consider a random variable experiment where we do Bernoulli trial independently with $P(success) = p$ until the first success occurs.

   (a) Construct a proper probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and define a proper random variable X so to induce a proper PMF for it.
   Solution:
   Given random experiment consists of sequential Benoulli subtrials which is independent and homogeneous with parameter p. Since a trial always has arbitrary $k(\geq 0)$ failures followed by last one success, then the natural sample space is

$$\Omega = \{S, FS, FFS, FFFS, \ldots, \underbrace{F\ldots F}_{k}S, \ldots\}, \tag{2.22}$$

   and we can find specific sigma algebra $\mathcal{A} \subset 2^{\Omega}$ satisfying following properties:

   - $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$
   - $\emptyset \in \mathcal{A}$ $(\Omega \in \mathcal{A})$
   - $E_1, E_2, E_3, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$.

   Also, for any single event E_k with k failures preceeding the first success, the probability measure $\mathbb{P}$ is defined by

$$\mathbb{P}(E_k) = \mathbb{P}[\{\underbrace{F\ldots F}_{k}S\}] = p(1-p)^k \quad \forall k \in \{0, 1, 2, 3, \ldots\} \tag{2.23}$$

   Therefore we now have the 'primitive' probability space $(\Omega, \mathcal{A}, \mathbb{P})$ for our random experiment.

   Now, let us define X be a random variable, which is given by

$$X = \text{The number of failure preceeding the first success} \tag{2.24}$$

   then, we may consider the new probability space $(\mathbb{R}, \mathcal{B}, P_X)$ where $\mathbb{R}$ is real number (which contains natural number $\mathbb{N}$ where the probability assumed actually), $\mathcal{B}$ is Borel sigma algebra, and P_X is probability distribution induced by X defined by

$$P_X(X \in B) = \mathbb{P}[\{\omega \in \Omega : \omega \in X^{-1}(B)\}]. \tag{2.25}$$

   Then, we can evaluate the probability mass function $f_X(x)$ from the probability distribution.

$$\begin{aligned} f_X(x) &= P_X(X = x) = P_X(X \in \{x\}) \\ &= \mathbb{P}[\{\omega \in \Omega : \omega \in X^{-1}(\{x\})\}] \\ &= \mathbb{P}[\{\underbrace{F\ldots F}_{x}S\}] \\ &= p(1-p)^x \quad \forall x \in \{0, 1, 2, 3, \ldots\} \end{aligned} \tag{2.26}$$

   ■

   (b) Compute the mean and variance of X.
   Solution:

mean: μ

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} x f_X(x) = \sum_{x=0}^{\infty} x p (1-p)^x \\
&= p(1-p) \sum_{x=0}^{\infty} x (1-p)^{x-1} \\
&= p(1-p) \frac{d}{dp} \left(- \sum_{x=0}^{\infty} (1-p)^x \right) \\
&= p(1-p) \frac{d}{dp} \left(1 - \frac{1}{p} \right) \\
&= p(1-p) \frac{1}{p^2} = \frac{1-p}{p}
\end{aligned} \tag{2.27}$$

variance: σ^2

$$E[(X-\mu)^2] = E[X(X-1)] + E[X] - \{E[X]\}^2 = E[X(X-1)] + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} \tag{2.28}$$

where

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) f_X(x) = \sum_{x=0}^{\infty} x(x-1) p (1-p)^x \\
&= p(1-p)^2 \sum_{x=0}^{\infty} x(x-1)(1-p)^{x-2} \\
&= p(1-p)^2 \frac{d^2}{dp^2} \left(\sum_{x=0}^{\infty} (1-p)^x \right) \\
&= p(1-p)^2 \frac{d^2}{dp^2} \left(\frac{1}{p} - 1 \right) \\
&= p(1-p)^2 \frac{2}{p^3} = \frac{2(1-p)^2}{p^2}
\end{aligned} \tag{2.29}$$

Hence,

$$E[(X-\mu)^2] = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2} \tag{2.30}$$

$\blacksquare$

(c) Prove that X satisfies the memoryless property: $P(X \geq s | X \geq t) = P(X \geq s-t), s > t$. **Solution:** Distribution function $F_X(x)$ is given by

$$F_X(x) = P_X(X \leq x) = \sum_{t=0}^{x} p(1-p)^t = 1 - (1-p)^{x+1} \tag{2.31}$$

Then,

$$\begin{aligned}
P(X \geq s | X \geq t) &= \frac{P(X \geq s, X \geq t)}{P(X \geq t)} = \frac{P(X \geq s)}{P(X \geq t)} \\
&= \frac{1 - P(X \leq s-1)}{1 - P(X \leq t-1)} = \frac{1 - F_X(s-1)}{1 - F_X(t-1)} \\
&= \frac{(1-p)^s}{(1-p)^t} = (1-p)^{s-t}
\end{aligned} \tag{2.32}$$

Then,

$$P(X \geq s-t) = 1 - P(X \leq s-t-1) = 1 - F_X(s-t-1) = (1-p)^{s-t} \tag{2.33}$$

Hence,

$$P(X \geq s | X \geq t) = P(X \geq s-t) \tag{2.34}$$

■

(d) Suppose that the probability is 0.001 that a light bulb will fail on any given day, then what is the probability that it will last at least 3 days.
Solution:
Let the bulb failure be the success event S, then $p = 0.001$. Let assume that the bulb failure event on any given day is independent with that events on other days, then any bulb's 'failure-or-not event' on a day can be viewed as an event of the Bernoulli trial. Because any bulb can not fail more than two times, and therefore the bulb always works well but fails on the last day. Hence, we can use Geometric Random Variable X given by

$$X = \text{The lifetime (day) of a light bulb, including the day of failure.} \tag{2.35}$$

Then, the probability mass function is given by

$$f_X(x) = p(1-p)^{x-1} \quad \forall x = 1, 2, 3, \ldots \tag{2.36}$$

and the distribution function is given by

$$F_X(x) = P(X \leq x) = \sum_{t=1}^{x} f_X(t) = 1 - (1-p)^x \tag{2.37}$$

Then, the probability that the bulb will last at least 3 days is

$$P(X \geq 3) = 1 - P(X \leq 4) = 1 - F_X(4) = (1-p)^4 = 0.999^4 \tag{2.38}$$

■

4. Let X be a random variable having the geometric distribution with $P(success) = p$.

   (a) Obtain the pdf for the random varaible Y defined as

$$Y = \frac{1}{X}$$

   Solution:
   The probability mass function of X is given by

$$f_X(x) = p(1-p)^{x-1} \quad \forall x = 1, 2, 3, \ldots \tag{2.39}$$

   and by definition

$$\begin{aligned} f_Y(y) &= P_Y(Y = y) = \mathbb{P}(\{Y = y\}) \\ &= \mathbb{P}(\{1/X = y\}) \\ &= \mathbb{P}(\{X = 1/y\}) \\ &= P_X(X = 1/y) = f_X(1/y) \\ &= p(1-p)^{\frac{1}{y}-1} \end{aligned} \tag{2.40}$$

   ■

   (b) Obtain $E[Y]$ and compare it to $E[X]$.
   Solution:
   $E[Y]$
   Let us take the sequence $(y_k)_{k\in\mathbb{N}}$ in $\mathbb{R}$, which is given by

$$y_k = \frac{1}{x_k} = \frac{1}{k} \quad \forall k \in \mathbb{N} \tag{2.41}$$

   then we can evalute the $E[Y]$ as follows

$$\begin{aligned} E[Y] &= \sum_{k\in\mathbb{N}} y_k f_Y(y_k) \\ &= \sum_{k\in\mathbb{N}} y_k p(1-p)^{1/y_k - 1} \\ &= \frac{p}{1-p} \sum_{k\in\mathbb{N}} \frac{1}{k}(1-p)^k \end{aligned} \tag{2.42}$$

   Since $(1-p)^k/k = \int -(1-p)^{k-1} dp + C'$, then

$$\begin{aligned} E[Y] &= \frac{p}{1-p} \sum_{k\in\mathbb{N}} \frac{1}{k}(1-p)^k \\ &= \frac{p}{1-p} \sum_{k\in\mathbb{N}} \left(\int -(1-p)^{k-1}\, dp + C \right) \\ &= \frac{p}{1-p} \left(\int -\sum_{k\in\mathbb{N}} (1-p)^{k-1}\, dp + C \right) \\ &= \frac{p}{1-p} \left(\int -\frac{1}{p}\, dp + C \right) \\ &= \frac{p}{1-p} (C - \ln p) \end{aligned} \tag{2.43}$$

Intuitively, if $p \to 1$, then $E[Y] \to 1$, then $C = 0$. Hence,

$$E[Y] = -\frac{p}{1-p} \ln p \tag{2.44}$$

Since $E[X] = (1-p)/p^2$, then

$$E[Y] = \frac{1}{E[X]} \left(\frac{1}{p} \ln \frac{1}{p} \right) \tag{2.45}$$

∎

5. Given a random variable X with pdf $f_X = 1$ for $0 < x < 1$ and any two points a_1, a_2 in the interval (0, 1), such that $a_1 < a_2$ and $a_1 + a_2 \leq b$,

   (a) show that
   $$P[a_1 < X < (a_1 + a_2)] = a_2.$$

   Solution:
   From the probability distribution function $f_X(x)$, we can find distribution function $F_X(x)$, which is given by
   $$F_X(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases} \tag{2.46}$$
   Since $F_X(x) = P[X < x] = P[X \leq x]$, then
   $$\begin{aligned} P[a_1 < X < (a_1 + a_2)] &= P[X < (a_1 + a_2)] - P[X \leq a_1] \\ &= F_X(a_1 + a_2) - F_X(a_1) \\ &= a_1 + a_2 - a_1 = a_2 \end{aligned} \tag{2.47}$$

   ■

   (b) In general, if $f(x)$ is uniform in the interval (a, b), and if $a \leq a_1$, $a_1 \leq a_2$, and $a_1 + a_2 \leq b$, show that
   $$P[a_1 < X < (a_1 + a_2)] = \frac{a_2}{b - a}.$$

   Solution: First, probability density $f(x)$ and distribution function $F(x)$ is given by
   $$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \tag{2.48}$$
   and
   $$F(x) = \begin{cases} 0, & x \leq a \\ \frac{1}{b-a}(x - a), & a < x < b \\ 1, & x \geq b \end{cases} \tag{2.49}$$
   Since $F(x) = P[X < x] = P[X \leq x]$, then
   $$\begin{aligned} P[a_1 < X < (a_1 + a_2)] &= P[X < (a_1 + a_2)] - P[X \leq a_1] \\ &= F(a_1 + a_2) - F(a_1) \\ &= \frac{a_1 + a_2 - a}{b - a} - \frac{a_1 - a}{b - a} = \frac{a_2}{b - a} \end{aligned} \tag{2.50}$$

   ■

6. Compute the varaince $Var(X)$ of a random variable X following binomial distribution.
Solution:
Let the parameter of binomial random variable is given by n and p, each of which indicate the total number of the Bernoulli trial and $P(success)$ of those trials, respectively. Then, PMF of binomial random variable X is given by

$$f_X(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}. \tag{2.51}$$

Then, first the mean μ is given by

$$\begin{aligned}
\mu = E[X] &= \sum_{x=0}^{n} x f_X(x;n,p) \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} np \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{(n-1)-x} \\
&= np \sum_{x=0}^{n-1} f_X(x;n-1,p) = np
\end{aligned} \tag{2.52}$$

and then, $E[X^2]$ is given by

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{n} x^2 f_X(x;n,p) = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= np \sum_{x=0}^{n-1} (x+1) \binom{n-1}{x} p^x (1-p)^{(n-1)-x} \\
&= np \left[\sum_{x=0}^{n-1} x f_X(x;n-1,p) + \sum_{x=0}^{n-1} f_X(x;n-1,p) \right] \\
&= np \left[np \sum_{x=0}^{n-1} f_X(x;n-2,p) + 1 \right] \\
&= np((n-1)p+1)
\end{aligned} \tag{2.53}$$

Then,

$$Var[X] = E[X^2] - \{E[X]\}^2 = np((n-1)p+1) - (np)^2 = np(1-p) = npq \tag{2.54}$$

■