Distribution of a function of a random variable

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 5.2

Aaron Osgood-Zimmerman

Department of Statistics

Practice solution

Practice

Let X be a discrete r.v. s.t.

$$\mathbb{P}(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad \text{for } k\in\{0,1,\dots\}$$

1. Check that $\sum_{k=0}^{+\infty}\mathbb{P}(X=k)=1$
2. Compute $\mathbb{E}[X], \mathrm{Var}[X]$ for $\lambda=1$

Reminder: The Taylor series expansion of exp is $e^x=\sum_{n=0}^{+\infty}\frac{x^k}{k!}$ for any $x\in\mathbb{R}$.

Solution

- First question is clear from the reminder
- Now

$$\mathbb{E}[X]=\sum_{k=0}^{+\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{+\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda\sum_{j=0}^{+\infty}\frac{\lambda^j}{j!}e^{-\lambda}=\lambda$$

- Similarly $\mathbb{E}[X(X-1)]=\lambda^2\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}=\lambda^2$
- Hence $\mathbb{E}[X^2]=\mathbb{E}[X(X-1)]+\mathbb{E}[X]=\lambda^2+\lambda$ and $\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\lambda$
- So for $\lambda=1$ we get $\mathbb{E}[X]=\mathrm{Var}[X]=1$.

Recap

Variance

- Variance
$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}[X])^2\right) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
- Standard deviation: $\sqrt{\mathrm{Var}(X)}$
- for $X \sim \mathrm{Ber}(p)$, $\mathrm{Var}(X) = p(1-p)$
- Property for $a, b \in \mathbb{R}$
$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$$
- $\mathrm{Var}(X) = 0 \iff \mathbb{P}(X = a) = 1$ for some $a \in \mathbb{R}$

Moments

- **nth moment** of a r.v. X is $\mathbb{E}[X^n]$
- **nth centered moment** of a r.v. X is $\mathbb{E}[(X-\mu)^n]$
- Expectation is the first moment
- Variance is the second centered moment

Outline

Function of a discrete r.v.

Exercise

Consider a game where you win with equal prob. $-1, 0, 1$ or 2\$.

What is the p.m.f. of the square of your gain?

Solution

- Let $X \sim \text{Unif}(\{-1, 0, 1, 2\})$ be your gain (uniform dist. on the discrete set $\{-1, 0, 1, 2\}$)
- Let $Y = X^2$ the square of your gain, note that $Y \in \{0, 1, 4\}$
- We can compute

$$\begin{aligned}
\mathbb{P}(Y = 0) &= \mathbb{P}(X^2 = 0) = \mathbb{P}(X = 0) = 1/4 \\
\mathbb{P}(Y = 1) &= \mathbb{P}(X^2 = 1) = \mathbb{P}(X = 1 \text{ or } X = -1)) \\
&= \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = 1/2 \\
\mathbb{P}(Y = 4) &= \mathbb{P}(X^2 = 4) = \mathbb{P}(X = 2 \text{ or } X = -2) \\
&= \mathbb{P}(X = 2) + \mathbb{P}(X = -2) = 1/4 + 0 = 1/4
\end{aligned}$$

Images and pre-images

Motivation

- Clearly getting the distribution of a function of a discrete r.v. is easy
- Let us introduce a few more definitions to clarify what happens

Definition (Images and pre-images)
Let E, F be two sets and $g : E \to F$. The **image** of $A \subseteq E$ by g is defined as

$$g(A) = \{g(x) : x \in A\} \subseteq F$$

The **pre-image** of $B \subseteq F$ is

$$g^{-1}(B) = \{x \in E : g(x) \in B\} \subseteq E$$

Note:

- Here g may not be invertible
- The above definition **applies on sets not on variables.**
- So the pre-image of a set always exists even if the inverse does not exist
- If there is no element that maps onto B, then $g^{-1}(B) = \emptyset$

Function of a discrete r.v.

Lemma
Let X be a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$, denote $Y = g(X)$.

The p.m.f. of Y is given as

$$p_Y(\ell) = \mathbb{P}(g(X) = \ell) = \mathbb{P}(X \in g^{-1}(\{\ell\})) = \sum_{k:g(k)=\ell} p_X(k)$$

Proof Follows from the definition of the pre-image, namely

$$g^{-1}(\{\ell\}) = \{k : g(k) = \ell\}$$

and $g(X) = \ell \iff X \in g^{-1}(\{\ell\})$.

Outline

Function of a discrete r.v.

Discrete function of a continuous r.v.

Continuous function of a continuous r.v.

Proofs

Distribution of a discrete function of a continuous r.v.

Exercise

Consider the score of a student to be uniformly distributed[1] on $X \sim \text{Unif}[0, 100]$.

A teacher rounds the scores of the students to the nearest integer (rounded up in case of two numbers equally distant).

What is the p.m.f. of the resulting scores?

Solution

- For $x \in \mathbb{R}$, let $g(x)$ be the nearest integer of x
- The variable of interest is Y that takes values in $g([0, 100]) = \{0, \ldots, 100\}$
- We have for $k \in \{1, \ldots, 99\}$

$$\mathbb{P}(Y = k) = \mathbb{P}(X \in [k-1/2, k+1/2)) = \int_{k-1/2}^{k+1/2} \frac{1}{100} dx = \frac{k+1/2}{100} - \frac{k-1/2}{100} = \frac{1}{100}$$

- And for $k = 0$ or $k = 100$,

$$\mathbb{P}(Y = 0) = \mathbb{P}(X \in [0, 1/2]) = \int_{0}^{1/2} \frac{1}{100} dx = \frac{1/2}{100}$$

$$\mathbb{P}(Y = 100) = \mathbb{P}(X \in [99 + 1/2, 100]) = \int_{99+1/2}^{100} \frac{1}{100} dx = \frac{1/2}{100}$$

Distribution of a discrete function of a continuous r.v.

More generally we have the following result

Lemma
Let X be a continuous r.v. and $g : \mathbb{R} \to \mathcal{Y}$ be a function that maps $\mathbb{R}$ onto a discrete set $\mathcal{Y}$.

Then the r.v. Y is discrete and for any $k \in \mathcal{Y}$,

$$\mathbb{P}(Y = k) = \mathbb{P}(X \in g^{-1}(\{k\}))$$

Note:

- We just retrieve the same property as before

Outline

Function of a discrete r.v.

Discrete function of a continuous r.v.

Continuous function of a continuous r.v.

Proofs

c.d.f. method

Method

- For continuous r.v., the p.d.f. has no interpretation as a prob. dist.
- It is easier to compute the c.d.f. of $Y = g(X)$ and then differentiate the c.d.f. to get the p.d.f.
- We will first illustrate the idea and then detail the method if
  - g is invertible
  - g is not invertible on $\mathbb{R}$ but invertible on some intervals of $\mathbb{R}$ that form a partition of $\mathbb{R}$

c.d.f. method

Exercise

Let X be a continuous r.v. with p.d.f. f_X

1. What is the p.d.f. of $Y = aX + b$ with $a > 0$?
2. Same but with $a < 0$?

Solution

1. - Compute the c.d.f. from the definition of the function,
     $$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(aX + b \leq t) = \mathbb{P}\left(X \leq \frac{t-b}{a}\right) = F_X\left(\frac{t-b}{a}\right)$$
   - Differentiate the c.d.f. to get the p.d.f.
     $$f_Y(y) = F_Y'(y) = \frac{1}{a} f_X\left(\frac{t-b}{a}\right)$$
2. - Similarly, for $a < 0$,
     $$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(aX + b \leq t) = \mathbb{P}\left(X \geq \frac{t-b}{a}\right) = 1 - F_X\left(\frac{t-b}{a}\right)$$
   - Therefore
     $$f_Y(y) = F_Y'(y) = -\frac{1}{a} f_X\left(\frac{t-b}{a}\right)$$
3. In any case we get $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right)$

c.d.f. method

Method for g invertible and strictly increasing

1. Compute

$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(X \leq g^{-1}(t)) = F_X(g^{-1}(t))$$

2. Deduce

$$\begin{aligned} f_Y(t) &= \frac{d}{dt} F_X(g^{-1}(t)) \\ &= (g^{-1}(t))' F'(g^{-1}(t)) \\ &= \frac{1}{g'(g^{-1}(t))} f_X(g^{-1}(t)) \end{aligned}$$

where we used that

- $(g^{-1}(t))' = 1/g'(g^{-1}(t))$
  (to remember that differentiate the equation $g(g^{-1}(x))) = x$ on both sides)
- and that $F_X'(z) = f_X(z)$

c.d.f. method

Method for g invertible and strictly decreasing

1. Compute

$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(X \geq g^{-1}(t)) = 1 - F_X(g^{-1}(t))$$

2. Deduce

$$f_Y(t) = -\frac{d}{dt} F_X(g^{-1}(t)) = -\frac{1}{g'(g^{-1}(t))} f_X(g^{-1}(t))$$

Distribution of a continuous function of a continuous r.v.

Previous considerations an be summarized by the following lemma

Lemma
Let X be a continuous r.v. with p.d.f. f_X

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and strictly monotonic

with inverse denoted $\gamma = g^{-1}$, then the p.d.f of $Y = g(X)$ exists[2] *and it reads*

$$f_Y(y) = \begin{cases} |\gamma'(y)| f_X(\gamma(y)) & \text{if } y \in g(\mathbb{R}) \\ 0 & \text{otherwise} \end{cases}$$

where $\gamma'(y) = \frac{1}{g'(g^{-1}(y))}$

Note:

- It is preferable to remember the method rather than the lemma because the method is more flexible (see next slides)

[2] We admit that fact

c.d.f. method

Details

- What if g is not defined on $\mathbb{R}$?
- The only thing you need is that g is defined on a subset $B \subseteq \mathbb{R}$ s.t.

$$\mathbb{P}(X \in B) = 1$$

- For $y \notin g(B)$, define $f_Y(y) = 0$

c.d.f. method

Exercise

Let $X \sim \text{Unif}([0,1])$ and $g : x \to -\frac{1}{\lambda}\log(1-x)$, where $\lambda > 0$.

What is the distribution of $Y = g(X)$?

Solution

- We simply need to look at the interval $(0,1)$ because $\mathbb{P}(X \in (0,1)) = \int_0^1 f_X(x)dx = 1$
- For $x \in (0,1)$, $g(x) = -\frac{1}{\lambda}\log(1-x) > 0$
- Therefore Y only takes positive values, so for $t \leq 0$ $\mathbb{P}(Y \leq t) = 0$
- Now for $t > 0$,

$$\begin{aligned}\mathbb{P}(Y \leq t) &= \mathbb{P}(-\frac{1}{\lambda}\log(1-X) \leq t) \\ &= \mathbb{P}(\log(1-X) \geq -\lambda t) = \mathbb{P}(1-X \geq e^{-\lambda t}) = \mathbb{P}(X \leq 1-e^{-\lambda t}) = 1-e^{-\lambda t}\end{aligned}$$

  because for $a \in [0,1]$, $\mathbb{P}(X \leq a) = \int_0^1 a dx = a$.
- So

$$f_Y(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Practice next lecture

Practice

Let $X \sim \mathrm{Exp}(\lambda)$, what is the p.d.f. of $Y = \sqrt{X}$?

Outline

Distribution of a continuous function of a continuous r.v.

Lemma

Let X be a continuous r.v. with p.d.f. f_X

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and strictly monotonic

with inverse denoted $\gamma = g^{-1}$, then the p.d.f of $Y = g(X)$ exists[3] *and it reads*

$$f_Y(y) = \begin{cases} |\gamma'(y)| f_X(\gamma(y)) & \text{if } y \in g(\mathbb{R}) \\ 0 & \text{otherwise} \end{cases}$$

where $\gamma'(y) = \frac{1}{g'(g^{-1}(y))}$

Proof Denote $a = \inf_x g(x)$, $b = \sup_x g(x)$, (potentially $a = -\infty$, $b = +\infty$)

1. If $t < a$, $F_Y(t) = \mathbb{P}(g(X) \leq t) = 0$ so $f_Y(t) = 0$
2. If $t < b$, $F_Y(t) = \mathbb{P}(g(X) \leq b) = 1$ so $f_Y(t) = 0$
3. Since the probability on a point does not matter we can define $f_Y(b) = f_Y(a) = 0$ if a, b are finite
4. if g is strictly increasing, for $t \in (a, b)$ s.t. $\gamma(t) := g^{-1}(t)$ is defined,
$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(g(X) \leq t) = \mathbb{P}(X \leq g^{-1}(t)) = F_X(\gamma(t))$$
$$\text{so} \quad f_Y(t) = \gamma'(t) f_X(\gamma(t))$$
5. if g is strictly decreasing, for $t \in (a, b)$ s.t. $g^{-1}(t)$ is defined,
$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(g(X) \leq t) = \mathbb{P}(X \geq g^{-1}(t)) = 1 - F_X(\gamma(t))$$
$$\text{so} \quad f_Y(t) = -\gamma'(t) f_X(\gamma(t))$$

Additional details

Pre-image

- Note that $\{g(X) \in A\} = \{X \in g^{-1}(A)\}$ and $g^{-1}(g(B)) \supseteq B$

Function of a r.v.

- Rigorously, Y is defined as a function $Y : \Omega \to \mathbb{R}$ that measures subsets in $\mathbb{R}$ as

$$\mu_Y : A \to \mathbb{P}(X \in g^{-1}(A)) \quad \text{for } A \text{ a Borel set in } \mathbb{R}$$

- This function μ_Y satisfies $0 \leq \mu_Y(A) \leq 1$
- Moreover for any two sets A, B, $g^{-1}(A \cup B) = g^1(A) \cup g^{-1}(B)$ and $g^{-1}(A \cap B) = g^1(A) \cap g^{-1}(B)$
- So μ_Y satisfies the additive property of a prob. measure
- For μ_Y to define a proper probability distribution we then need

$$1 = \mathbb{P}(X \in g^{-1}(\mathbb{R}))$$

  which is the case if $g^{-1}(\mathbb{R}) \supseteq B$ where $B \subseteq \mathbb{R}$ is s.t. $\mathbb{P}(X \in B) = 1$, i.e., g is defined on B