Statistical Inference

Lecture 2: Transformations and Expectations

MING GAO

DASE @ ECNU
(for course related communications)
mgao@dase.ecnu.edu.cn

Mar. 8, 2018

Outline

Outline

Functions of a r.v.

If X is a r.v., then any function of X, say $g(X)$, is also a r.v..

Functions of a r.v.

If X is a r.v., then any function of X, say $g(X)$, is also a r.v..

- A natural question is whether we can describe the probabilistic behavior of Y in terms of that of X.

Functions of a r.v.

If X is a r.v., then any function of X, say $g(X)$, is also a r.v..

- A natural question is whether we can describe the probabilistic behavior of Y in terms of that of X.
- That is, for any set A,

$$P(Y \in A) = P(g(X) \in A).$$

Functions of a r.v.

If X is a r.v., then any function of X, say $g(X)$, is also a r.v..

- A natural question is whether we can describe the probabilistic behavior of Y in terms of that of X.
- That is, for any set A,

$$P(Y \in A) = P(g(X) \in A).$$

- We associate with g an inverse mapping, denoted as g^{-1}, which is a mapping from subsets of Y to subsets of X, and is defined by $g^{-1}(A) = \{x \in X | g(x) \in A\}$.

Functions of a r.v.

If X is a r.v., then any function of X, say $g(X)$, is also a r.v..

- A natural question is whether we can describe the probabilistic behavior of Y in terms of that of X.
- That is, for any set A,

$$P(Y \in A) = P(g(X) \in A).$$

- We associate with g an inverse mapping, denoted as g^{-1}, which is a mapping from subsets of Y to subsets of X, and is defined by $g^{-1}(A) = \{x \in X | g(x) \in A\}$.
- For any set $A \subset \mathcal{Y}$,

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A)).$$

Example of discrete transformation

Binomial transformation

A discrete r.v. X has a binomial distribution if its pmf is of the form

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \cdots, n.$$

Example of discrete transformation

Binomial transformation

A discrete r.v. X has a binomial distribution if its pmf is of the form

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \cdots, n.$$

Consider the r.v. $Y = g(X) = n - X$, and $\mathcal{Y} = \{y | y = g(x), x \in \mathcal{X}\} = \{0, 1, \cdots, n\}$.

Example of discrete transformation

Binomial transformation

A discrete r.v. X has a binomial distribution if its pmf is of the form

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \cdots, n.$$

Consider the r.v. $Y = g(X) = n - X$, and $\mathcal{Y} = \{y | y = g(x), x \in \mathcal{X}\} = \{0, 1, \cdots, n\}$.

$$\begin{aligned} f_Y(y) &= \sum_{x \in g^{-1}(y)} f_X(x) = f_X(n-y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} = \binom{n}{y} (1-p)^y p^{n-y}. \end{aligned}$$

Example of continuous transformation

Uniform transformation

Suppose X has a uniform distribution on the interval $(0, 2\pi)$, that is

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 < x < 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

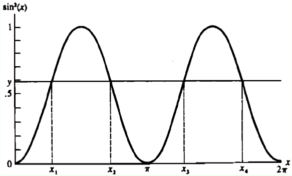
Example of continuous transformation

Uniform transformation

Suppose X has a uniform distribution on the interval $(0, 2\pi)$, that is

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 < x < 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

Consider the r.v. $Y = \sin^2(X)$, and $\mathcal{Y} \in [0, 1]$.



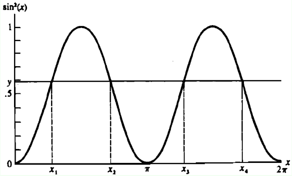
Example of continuous transformation

Uniform transformation

Suppose X has a uniform distribution on the interval $(0, 2\pi)$, that is

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 < x < 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

Consider the r.v. $Y = \sin^2(X)$, and $\mathcal{Y} \in [0, 1]$.



$$\begin{aligned} f_Y(y) &= P(Y \leq y) \\ &= P(X \leq x_1) + P(x_2 \leq X \leq x_3) + P(X \geq x_4) \\ &= 2P(X \leq x_1) + 2P(x_2 \leq X \leq \pi), \end{aligned}$$

where x_1 and x_2 are the two solutions to $\sin^2(x) = y$ for $0 < x < \pi$.

Outline

Theorem for monotone transformations

Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x | f_X(x) > 0\} \text{ and } \mathcal{Y} = \{y | y = g(x) \text{ for } x \in \mathcal{X}\}.$$

Theorem for monotone transformations

Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x | f_X(x) > 0\} \text{ and } \mathcal{Y} = \{y | y = g(x) \text{ for } x \in \mathcal{X}\}.$$

a. If g is increasing on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$;

Theorem for monotone transformations

Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x | f_X(x) > 0\} \text{ and } \mathcal{Y} = \{y | y = g(x) \text{ for } x \in \mathcal{X}\}.$$

a. If g is increasing on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$;
b. If g is decreasing on $\mathcal{X}$ and X is a continuous r.v., $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$;

Theorem for monotone transformations

Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x|f_X(x) > 0\} \text{ and } \mathcal{Y} = \{y|y = g(x) \text{ for } x \in \mathcal{X}\}.$$

a. If g is increasing on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$;
b. If g is decreasing on $\mathcal{X}$ and X is a continuous r.v., $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$;

Proof.
[a.] $\{x \in \mathcal{X}|g(x) \leq y\} = \{x \in \mathcal{X}|x \leq g^{-1}(y)\}$ since If g is increasing. Furthermore, we have

$$F_Y(y) = \int_{x \in \mathcal{X}: x \leq g^{-1}(y)} f_X(x)dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X(g^{-1}(y)).$$

Uniform exponential relationship

Suppose $X \sim f_X(x) = 1$ if $0 < x < 1$ and 0 otherwise, the uniform $(0, 1)$ distribution. It is straightforward to check that $F_X(x) = x, 0 < x < 1$. Let $Y = g(X) = -\log X$.

$$\frac{d}{dx} g(x) = \frac{d}{dx}(-\log x) = \frac{-1}{x} < 0, \text{ for } 0 < x < 1.$$

Uniform exponential relationship

Suppose $X \sim f_X(x) = 1$ if $0 < x < 1$ and 0 otherwise, the uniform $(0,1)$ distribution. It is straightforward to check that $F_X(x) = x, 0 < x < 1$. Let $Y = g(X) = -\log X$.

$$\frac{d}{dx}g(x) = \frac{d}{dx}(-\log x) = \frac{-1}{x} < 0, \text{ for } 0 < x < 1.$$

That is $g(x)$ is a decreasing function. As $X \in [0,1]$ and $-\log x \in [0, \infty]$.

Uniform exponential relationship

Suppose $X \sim f_X(x) = 1$ if $0 < x < 1$ and 0 otherwise, the uniform $(0, 1)$ distribution. It is straightforward to check that $F_X(x) = x, 0 < x < 1$. Let $Y = g(X) = -\log X$.

$$\frac{d}{dx}g(x) = \frac{d}{dx}(-\log x) = \frac{-1}{x} < 0, \text{ for } 0 < x < 1.$$

That is $g(x)$ is a decreasing function. As $X \in [0, 1]$ and $-\log x \in [0, \infty]$.

For $y > 0$, $y = -\log x$ implies $x = e^{-y}$, therefore,

$$F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}.$$

Of course, $F_Y(y) = 0$ for $y \leq 0$.

Theorem for continuous r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dx}g^{-1}(y)|, & y \in \mathcal{Y}; \\ 0, & \textit{otherwise}. \end{cases}$$

Proof.

Theorem for continuous r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dx}g^{-1}(y)|, & y \in \mathcal{Y}; \\ 0, & otherwise. \end{cases}$$

Proof.

By the chain rule,

$$f_Y(y) = \frac{d}{dx}F_Y(y) = \begin{cases} f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y), & \text{if g is increasing;} \\ -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y), & \text{if g is decreasing.} \end{cases}$$

□

Example

Inverted Gamma pdf

Question: Suppose X has the Gamma pdf

$$f(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, 0 < x < \infty.$$

Suppose we want to find the pdf of $g(X) = \frac{1}{X}$.

Example

Inverted Gamma pdf

Question: Suppose X has the Gamma pdf

$$f(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, 0 < x < \infty.$$

Suppose we want to find the pdf of $g(X) = \frac{1}{X}$.
Solution:
If we let $y = g(x)$, then $g^{-1}(y) = \frac{1}{y}$ and $\frac{d}{dy} g^{-1}(y) = \frac{-1}{y^2}$.

Example

Inverted Gamma pdf

Question: Suppose X has the Gamma pdf

$$f(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, 0 < x < \infty.$$

Suppose we want to find the pdf of $g(X) = \frac{1}{X}$.
Solution:
If we let $y = g(x)$, then $g^{-1}(y) = \frac{1}{y}$ and $\frac{d}{dy} g^{-1}(y) = \frac{-1}{y^2}$.
Applying the above theorem, for $y \in (0, \infty)$, we get

$$\begin{aligned} f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dx} g^{-1}(y)| &= \frac{1}{(n-1)!\beta^n} (\frac{1}{y})^{n-1} e^{-1/y\beta} \frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n} (\frac{1}{y})^{n+1} e^{-1/y\beta}. \end{aligned}$$

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \leq i \leq k$.

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \leq i \leq k$.

- $g(x) = g_i(x)$, for $x \in A_i$;

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \leq i \leq k$.

- $g(x) = g_i(x)$, for $x \in A_i$;
- $g_i(x)$ is monotone on A_i;

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \leq i \leq k$.

- $g(x) = g_i(x)$, for $x \in A_i$;
- $g_i(x)$ is monotone on A_i;
- the set $\mathcal{Y} = \{y | y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $1 \leq i \leq k$;

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \leq i \leq k$.

- $g(x) = g_i(x)$, for $x \in A_i$;
- $g_i(x)$ is monotone on A_i;
- the set $\mathcal{Y} = \{y | y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $1 \leq i \leq k$;
- $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$ for $1 \leq i \leq k$.

Theorem for non-monotone r.v.

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \cdots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_i(x)$ defined on A_i for $1 \le i \le k$.

- $g(x) = g_i(x)$, for $x \in A_i$;
- $g_i(x)$ is monotone on A_i;
- the set $\mathcal{Y} = \{y | y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $1 \le i \le k$;
- $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$ for $1 \le i \le k$.

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) |\frac{d}{dx} g_i^{-1}(y)|, & y \in \mathcal{Y}; \\ 0, & otherwise. \end{cases}$$

Normal-Chi squared relationship

Suppose X has the standard normal distribution,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, -\infty < x < +\infty.$$

Suppose $Y = g(X) = X^2$. Since $g(x) = x^2$ is monotone on $(-\infty, 0)$ and $(0, +\infty)$. Applying above theorem, we take

$$A_0 = \{0\}; A_1 = (-\infty, 0), g_1(x) = x^2, g_1^{-1}(y) = -\sqrt{y};$$
$$A_2 = (0, +\infty), g_2(x) = x^2, g_2^{-1}(y) = \sqrt{y}.$$

Normal-Chi squared relationship

Suppose X has the standard normal distribution,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, -\infty < x < +\infty.$$

Suppose $Y = g(X) = X^2$. Since $g(x) = x^2$ is monotone on $(-\infty, 0)$ and $(0, +\infty)$. Applying above theorem, we take

$$\begin{aligned} A_0 = \{0\}; A_1 = (-\infty, 0), g_1(x) = x^2, g_1^{-1}(y) = -\sqrt{y}; \\ A_2 = (0, +\infty), g_2(x) = x^2, g_2^{-1}(y) = \sqrt{y}. \end{aligned}$$

Then the pdf of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, 0 < y < \infty. \end{aligned}$$

Probability integral transformation

Let X have continuous cdf $F_X(x)$ and define the r.v. Y as $Y = F_X(X)$. Then Y is uniformly distributed on $[0, 1]$, that is, $P(Y \leq y) = y, 0 < y < 1$.

Probability integral transformation

Let X have continuous cdf $F_X(x)$ and define the r.v. Y as $Y = F_X(X)$. Then Y is uniformly distributed on $[0, 1]$, that is, $P(Y \leq y) = y, 0 < y < 1$.

Proof.

For $Y = F_X(X)$ and $0 < y < 1$, we have

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

$\square$

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

- After reading the fine print. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?
- If 10,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?
- If 10,000 tickets will be sold. Is this a good bet?
- If 100,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.
If there are 10,000 tickets will be sold, how about your answer?

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

Enter the Charity Lottery
One Grand Prize of \$20,000
20 additional prizes of \$500

Tickets only \$10

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.
If there are 10,000 tickets will be sold, how about your answer?

Expected value

Definition

The **expected value or mean** of a r.v. $g(X)$, denoted as $E(g(X))$, is

$$E(g(X)) = \begin{cases} \int_{-\infty}^{+\infty} g(x) f_X(x) dx, & \text{if X is continuous;} \\ \sum_{x \in X} g(x) P(X = x), & \text{if X is discrete;} \end{cases}$$

Expected value

Definition

The **expected value or mean** of a r.v. $g(X)$, denoted as $E(g(X))$, is

$$E(g(X)) = \begin{cases} \int_{-\infty}^{+\infty} g(x) f_X(x) dx, & \text{if X is continuous;} \\ \sum_{x \in X} g(x) P(X = x), & \text{if X is discrete;} \end{cases}$$

- The **deviation** of X at $\omega \in \Omega$ is $X(\omega) - E(X)$, the difference between the value of X and the mean of X.
- If $E|g(X)| = \infty$, we say that $E(g(X))$ does not exist.

Example I

Exponential mean

Question: Suppose X has an exponential (λ) distribution,

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, 0 \leq x < +\infty, \lambda > 0.$$

Suppose $Y = g(X) = X^2$.
Solution:
The $E(X)$ is given by

$$\begin{aligned} E(X) &= \int_0^\infty \frac{1}{\lambda} x e^{-x/\lambda} dx = -xe^{-x/\lambda}|_0^\infty + \int_0^\infty e^{-x/\lambda} dx \\ &= \int_0^\infty e^{-x/\lambda} dx = \lambda. \end{aligned}$$

Example II

Binomial mean

Question: Suppose X has a binomial distribution, its pmf is given by

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,\cdots,n.$$

Solution:

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} x\cdot P(X=x)=\sum_{x=1}^{n} x\cdot\binom{n}{x}p^x(1-p)^{n-x}\\
&= \sum_{x=1}^{n} n\cdot\binom{n-1}{x-1}p^x q^{n-x}=np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j q^{n-1-j}\\
&= np(p+q)^{n-1}=np
\end{aligned}$$

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Proof.

We have known that $P(X = k) = q^{k-1}p$. Hence, we have

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \cdot q^{k-1} p = p(\sum_{m=1}^{\infty} \sum_{k=m}^{\infty} q^{k-1}) \\
&= p(\sum_{m=1}^{\infty} \frac{q^{m-1}}{1-q}) = \sum_{m=1}^{\infty} q^{m-1} \\
&= \frac{1}{1-q} = \frac{1}{p}
\end{aligned}
$$

□

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Proof.

We have known that $P(X = k) = q^{k-1}p$. Hence, we have

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot q^{k-1} p = p(\sum_{m=1}^{\infty} \sum_{k=m}^{\infty} q^{k-1}) \\ &= p(\sum_{m=1}^{\infty} \frac{q^{m-1}}{1-q}) = \sum_{m=1}^{\infty} q^{m-1} \\ &= \frac{1}{1-q} = \frac{1}{p} \end{aligned}$$

□

Cauchy mean

A classic example of a r.v. whose expected value does not exist is a Cauchy r.v., that is, one with pdf

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, -\infty < x < \infty.$$

It is straightforward to check that $\int_{-\infty}^{+\infty} f_X(x)dx = 1$, but $E|X| = \infty$. For any positive number M,

$$\begin{aligned} E|X| &= \int_{-\infty}^{+\infty} \frac{|x|}{\pi}\frac{1}{1+x^2}dx = \frac{2}{\pi}\int_0^{+\infty}\frac{x}{1+x^2}dx \\ &= \lim_{M\to\infty}\frac{2}{\pi}\int_0^{M}\frac{x}{1+x^2}dx = \frac{1}{\pi}\lim_{M\to\infty}\log\left(1+M^2\right) = \infty \end{aligned}$$

and $E(X)$ does not exist.

Outline

Linearity of expectations

Theorem

Let X be a r.v. and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

a. $E(ag_1(X) + bg_2(X) + c) = aE(g_1(X)) + bE(g_2(X)) + c$;
b. If $g_1(x) \geq 0$ for all x, then $E(g_1(X)) \geq 0$;
c. If $g_1(x) \geq g_2(x)$ for all x, then $E(g_1(X)) \geq E(g_2(X))$;
d. If $a \leq g_1(x) \leq b$ for all x, then $a \leq E(g_1(X)) \leq b$;

Expected value of Bernoulli trials

Proof with linearity of expectations

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Expected value of Bernoulli trials

Proof with linearity of expectations

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Proof.

Let X_i be # heads in the $i-$th Bernoulli trial, and X be the number of successes in n mutually independent Bernoulli trials. Hence we have $X = \sum_{i=1}^{n} X_i$, and $E(X_i) = p$.

$$E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = np.$$

□

Uniform exponential relationship

Suppose $X \sim f_X(x) = 1$ if $0 \leq x \leq 1$ and 0 otherwise, the uniform $(0, 1)$ distribution. Let $Y = g(X) = -\log X$.

$$E(g(X)) = E(-\log X) = \int_0^1 -\log x dx = x - x \log x|_0^1 = 1.$$

We also have $Y = -\log X$ has cdf $1 - e^{-y}$, and pdf $f_Y(y) = \frac{d}{dy}(1 - e^{-y}) = e^{-y}, 0 < y < \infty$, which is a special case of the exponential pdf with $\lambda = 1$. Thus, $E(Y) = 1$.

Outline

Moment

For each integer n, the $n-$th moment of X, $\mu_n^{'}$, is $\mu_n^{'} = E(X^n)$. The $n-$th central moment of X, μ_n, is $\mu_n = E(X - \mu)^n$, where $\mu = \mu_1^{'} = E(X)$.

Moment

For each integer n, the $n-$th moment of X, $\mu_n^{'}$, is $\mu_n^{'} = E(X^n)$. The $n-$th central moment of X, μ_n, is $\mu_n = E(X-\mu)^n$, where $\mu = \mu_1^{'} = E(X)$.

Variance

The variance of a r.v. X is its second central moment,

$$Var(X) = E(X-\mu)^2.$$

The positive square root of $Var(X)$ is the standard deviation of X.

- $Var(X) = E(X^2) - (E(X))^2$;
- The variance gives a measure of the degree of spread of a distribution around its mean.

Exponential variance

Let X have the *exponential*(λ) distribution. We can now calculate the variance by

$$
\begin{aligned}
Var(X) = E(X-\lambda)^2 &= \int_0^\infty (x-\lambda)^2 \frac{1}{\lambda} e^{-x/\lambda} dx \\
&= \int_0^\infty (x^2 - 2x\lambda + \lambda^2) \frac{1}{\lambda} e^{-x/\lambda} dx
\end{aligned}
$$

To complete the integration, we can integrate each of the terms separately, using integration by parts on the terms involving x and x^2.
Upon doing this, we find that

$$Var(X) = \lambda^2.$$

Variance of Bernoulli trial

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome $\#$ heads in this outcome. What is the variance of X if it is a biased coin with $P(\{H\}) = p$?

Variance of Bernoulli trial

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome # heads in this outcome. What is the variance of X if it is a biased coin with $P(\{H\}) = p$?

Solution:

$$E(X^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$
$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$
$$V(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$$

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?
Solution:

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \cdot P(X=k) = \sum_{k=1}^{n} k(k-1) \cdot P(X=k) + \sum_{k=1}^{n} k \cdot P(X=k) \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k} + np \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^{j} q^{n-2-j} + np \\
&= n(n-1)p^2 (p+q)^{n-2} + np = n(n-1)p^2 + np, \\
V(X) &= E(X^2) - (E(X))^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}
$$

Nonlinearity of variance

Theorem

If X is a r.v. on Ω, and if a and b are real numbers, then

$$V(aX+b) = a^2 V(X).$$

Proof.

$$\begin{aligned}
V(aX+b) &= E((aX+b)^2) - (E(aX+b))^2 \\
&= E((a^2X^2 + 2abX + b^2)) - (a^2(E(X))^2 + 2abE(X) + b^2) \\
&= a^2E(X^2) + 2abE(X) + b^2 - a^2(E(X))^2 - 2abE(X) - b^2 \\
&= a^2E(X^2) - a^2(E(X))^2 = a^2V(X).
\end{aligned}$$

$\square$

Bienaymé's formula

Theorem

Question: If X and Y are two independent r.v.s on a sample space Ω, then $V(X+Y) = V(X) + V(Y)$. Furthermore, if X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are pairwise independent r.v.s on Ω, then

$$V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i).$$

Proof:

$$\begin{aligned}
V(X+Y) &= E((X+Y)^2) - [E(X+Y)]^2. \\
&= E(X^2 + 2XY + Y^2) - ([E(X)]^2 + 2E(X)E(Y) + [E(Y)]^2) \\
&= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\
&= V(X) + V(Y)
\end{aligned}$$

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Solution:

Let X_i be the number of success in the $i-$th Bernoulli trial. Thus, we have $X = \sum_{i=1}^{n} X_i$, X_i and X_j are independent for $i \neq j$.

$$E(X_i^2) = 1^2 p + 0^2(1-p) = p;$$

$$V(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1-p);$$

$$V(X) = \sum_{i=1}^{n} V(X_i) = np(1-p).$$

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^{n} X_i$?

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^{n} X_i$?

Solution:

$$\begin{aligned} V(\frac{1}{n}\sum_{i=1}^{n} X_i) &= (\frac{1}{n})^2 V(\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 \sum_{i=1}^{n} V(X_i) \\ &= (\frac{1}{n})^2 n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^{n} X_i$?

Solution:

$$V(\frac{1}{n}\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 V(\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 \sum_{i=1}^{n} V(X_i)$$
$$= (\frac{1}{n})^2 n\sigma^2 = \frac{\sigma^2}{n}$$

That is, the variance of the mean decreases when n increases. It is a good property of variance.

Expectation of independent r.v.s

Theorem

If X and Y are independent r.v.s on a sample space Ω, then

$$E(XY) = E(X)E(Y).$$

Proof: To prove this formula, we use the key observation that event $XY = r$ is the disjoint union of events $X = r_1$ and $Y = r_2$ over all $r_1 \in X(\Omega)$ and $r_2 \in Y(\Omega)$ with $r = r_1 r_2$. We have

$$E(XY) = \sum_{r \in XY(\Omega)} r \cdot P(XY = r) = \sum_{r_1 \in X(\Omega), r_2 \in Y(\Omega)} r_1 r_2 \cdot P(X = r_1 \wedge Y = r_2)$$

$$= \sum_{r_1 \in X(\Omega)} \sum_{r_2 \in Y(\Omega)} (r_1 \cdot P(X = r_1))(r_2 \cdot P(Y = r_2))$$

$$= \sum_{r_1 \in X(\Omega)} (r_1 \cdot P(X = r_1) \sum_{r_2 \in Y(\Omega)} (r_2 \cdot P(Y = r_2))) = \sum_{r_1 \in X(\Omega)} (r_1 \cdot P(X = r_1) E(Y))$$

$$= E(Y) \sum_{r_1 \in X(\Omega)} r_1 \cdot P(X = r_1) = E(X)E(Y).$$

Outline

Moment generating functions

Definition

Let X be a r.v. with cdf $F_X(x)$. The moment generating function of X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = E(e^{tX}).$$

- The expectation exists for t in some neighborhood of 0 if there is an $h > 0$, such that $E(e^{tX})$ for all $t \in (-h, h)$.
- The moment generating function does not exist if the expectation does not exist.
- The moment generating function can be computed as
  - X is continuous, $M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx$;
  - X is discrete, $M_X(t) = \sum_x e^{tx} P(X = x)$.

Theorem

If X has moment generating function $M_X(t)$, then

$$E(X^n) = M_X^{(n)}(0),$$

where we define

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)|_{t=0}.$$

Proof.

$$\frac{d}{dt}M_X(t) = \frac{d}{dt}\int_{-\infty}^{+\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{+\infty} \frac{d}{dt}(e^{tx}) f_X(x)dx$$
$$= \int_{-\infty}^{+\infty} (xe^{tx}) f_X(x)dx = E(Xe^{tX}).$$

Proof Cont'd

Proof.

Thus,

$$\frac{d}{dt} M_X(t)|_{t=0} = E(Xe^{tX})|_{t=0} = EX.$$

Proceeding in an analogous manner, we can establish that

$$\frac{d^n}{dt^n} M_X(t)|_{t=0} = E(X^n e^{tX})|_{t=0} = EX^n.$$

$\square$

That is, the $n-$th moment is equal to the $n-$th derivative of $M_X(t)$ evaluated at $t = 0$.

Gamma moment generating function

The Gamma pdf is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx, 0 < x < \infty, \alpha > 0, \beta > 0.$$

The moment generating function is given by

$$\begin{aligned} M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} e^{tx} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-x/(\frac{\beta}{1-\beta t})} dx \end{aligned}$$

Note that $\int_0^{\infty} f_X(x)dx = 1$, hence $\int_0^{\infty} x^{\alpha-1} e^{-x/\beta} dx = \Gamma(\alpha)\beta^{\alpha}$.

Gamma moment generating function Cont'd

If $t < \frac{1}{\beta}$,

$$\begin{aligned} M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)\Big(\frac{\beta}{1-\beta t}\Big)^\alpha \\ &= \Big(\frac{1}{1-\beta t}\Big)^\alpha. \end{aligned}$$

The mean of the Gamma distribution is given by

$$E(X) = \frac{d}{dt}M_X(t)\Big|_{t=0} = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}\Big|_{t=0} = \alpha\beta.$$

Moment generating functions VS. distributions

Theorem

Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist

- If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $E(X^r) = E(Y^r)$ for all integers $r = 0, 1, 2, \cdots$
- If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.

Convergence of moment generating functions

Suppose $\{X_i, i = 1, 2, \cdots\}$ is a sequence of r.v., each with mgf $M_{X_i}(t)$. Furthermore, suppose that, for all t in a neighborhood of 0,

$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t),$$

where $M_X(t)$ is an mgf. Then there is a unique cdf $F_X(x)$ whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i \to \infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t| < h$, of mgfs to an mgf implies convergence of cdfs.

Property of moment generating functions

For any constants a and b, the mgf of the r.v. $aX + b$ is given by

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

Proof.

$$\begin{aligned} M_{aX+b}(t) &= E\left(e_{(aX+b)t}\right) \\ &= E\left(e_{(aX)t} e^{bt}\right) \\ &= e^{bt} E\left(e_{(at)X}\right) \\ &= e^{bt} M_X(at). \end{aligned}$$

$\square$

Leibnitz's Rule

If $f(x,\theta), a(\theta)$ and $b(\theta)$ are differentiable w.r.t. θ, then

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\frac{db(\theta)}{d\theta} - f(a(\theta),\theta)\frac{da(\theta)}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial f(x,\theta)}{\partial\theta}dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

$$\frac{d}{d\theta}\int_{a}^{b} f(x,\theta)dx = \int_{a}^{b} \frac{\partial f(x,\theta)}{\partial\theta}dx.$$

- If we have the integral of a differentiable function over a finite range, differentiation of the integral poses no problem;
- If the range of integration is infinite, problem can arise.

Interchanging differentiation and integration

Recall that if $f(x,\theta)$ is differentiable, then

$$\frac{\partial f(x,\theta)}{\partial \theta} = \lim_{\delta \to 0} \frac{f(x,\theta+\delta) - f(x,\theta)}{\delta},$$

So we have

$$\int_{-\infty}^{+\infty} \frac{\partial f(x,\theta)}{\partial \theta} dx = \int_{-\infty}^{+\infty} \lim_{\delta \to 0} \Big[\frac{f(x,\theta+\delta) - f(x,\theta)}{\delta}\Big] dx,$$

while

$$\frac{d}{d\theta} \int_{-\infty}^{+\infty} f(x,\theta) dx = \lim_{\delta \to 0} \int_{-\infty}^{+\infty} \Big[\frac{f(x,\theta+\delta) - f(x,\theta)}{\delta}\Big] dx.$$

Corollary of Lebesgue's Dominated Convergence Theorem

Suppose the function $h(x, y)$ is continuous at y_0 for each x, and there exists a function $g(x)$ satisfying

- $|h(x, y)| \leq g(x)$ for all x and y;
- $\int_{-\infty}^{+\infty} g(x)dx < \infty$.

Then,

$$\lim_{y \to y_0} \int_{-\infty}^{+\infty} h(x, y)dx = \int_{-\infty}^{+\infty} \lim_{y \to y_0} h(x, y)dx.$$

The key condition in this theorem is the existence of a dominating function $g(x)$, with a finite integral.

Corollary

Suppose the function $f(x,\theta)$ is continuous at θ_0, that is,

$$\lim_{\delta\to 0}\frac{f(x,\theta_0+\delta)-f(x,\theta_0)}{\delta}=\frac{\partial f(x,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}$$

exists for every x, and there exists a function $g(x,\theta_0)$ and a constant $\delta_0>0$ such that

- $\left|\frac{f(x,\theta_0+\delta)-f(x,\theta_0)}{\delta}\right|\le g(x,\theta_0)$ for all x and $|\delta|\le\delta_0$;
- $\int_{-\infty}^{+\infty} g(x,\theta_0)dx<\infty$.

Then,

$$\frac{d}{d\theta}\int_{-\infty}^{+\infty} f(x,\theta)dx\Big|_{\theta=\theta_0}=\int_{-\infty}^{+\infty}\Big[\frac{\partial f(x,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\Big]dx.$$

The distinction between θ and θ_0 is not stressed when $f(x,\theta)$ is differentiable at all θ.

Example

Let X have the *exponential*(λ) pdf given by $f(x) = \frac{1}{\lambda}e^{-x/\lambda}, 0 < x < \infty$, and suppose we want to calculate

$$\frac{d}{d\lambda}E(X^n) = \frac{d}{d\lambda}\int_0^\infty x^n \frac{1}{\lambda}e^{-x/\lambda}dx$$

for integer $n > 0$.
If we could move the differentiation inside the integral, we would have

$$\begin{aligned}\frac{d}{d\lambda}E(X^n) &= \int_0^\infty \frac{\partial}{\partial\lambda}x^n \frac{1}{\lambda}e^{-x/\lambda}dx \\ &= \int_0^\infty \frac{x^n}{\lambda^2}(\frac{x}{\lambda})e^{-x/\lambda}dx = \frac{1}{\lambda^2}E(X^{n+1}) - \frac{1}{\lambda}E(X^n).\end{aligned}$$

$$E(X^{n+1}) = \lambda E(X^n) + \lambda^2\frac{d}{d\lambda}E(X^n).$$

Justifying the interchange

We bound the derivative of $x^n \frac{1}{\lambda} e^{-x/\lambda}$. Now

$$|\frac{\partial}{\partial\lambda}(\frac{x^n e^{-x/\lambda}}{\lambda})| = \frac{x^n e^{-x/\lambda}}{\lambda^2}|\frac{x}{\lambda} - 1| \leq \frac{x^n e^{-x/\lambda}}{\lambda^2}(\frac{x}{\lambda} + 1).(\frac{x}{\lambda} > 0)$$

For $0 < \delta_0 < \lambda$ and $|\lambda^{'} - \lambda| \leq \delta_0$, we take

$$|\frac{\partial}{\partial\lambda}(\frac{x^n e^{-x/\lambda}}{\lambda})\big|_{\lambda=\lambda'}| \leq \frac{x^n e^{-x/(\lambda+\delta_0)}}{(\lambda-\delta_0)^2}(\frac{x}{\lambda-\delta_0} + 1) = g(x, \lambda).$$

Since the exponential distribution has all of its moments, i.e., $\int_{-\infty}^{+\infty} g(x, \lambda) < \infty$ as long as $\lambda - \delta_0 > 0$.

Thus, the example gives us a recursion relation for the moment of exponential distribution.

Interchanging differentiation and summation

Suppose that the series $\sum_{x=0}^{\infty} h(\theta, x)$ converges for all θ in an interval (a, b) of real numbers and

- $\frac{\partial}{\partial\theta} h(\theta, x)$ is continuous in θ for each x;
- $\sum_{x=0}^{\infty} \frac{\partial}{\partial\theta} h(\theta, x)$ converges uniformly on every closed bounded subinterval of (a, b).

Then,

$$\frac{d}{d\theta} \sum_{x=0}^{\infty} h(\theta, x) = \sum_{x=0}^{\infty} \frac{\partial}{\partial\theta} h(\theta, x).$$

The key condition in this theorem is the existence of a dominating function $g(x)$, with a finite integral.

Example

Let X be a geometric distribution $P(X=x)=\theta(1-\theta)^x, x=0,1,\cdots,0<\theta<1$. We have that $\sum_{x=0}^{\infty}\theta(1-\theta)^x=1$, and

$$\begin{aligned}\frac{d}{d\theta}\sum_{x=0}^{\infty}\theta(1-\theta)^x &= \sum_{x=0}^{\infty}\frac{d}{d\theta}\theta(1-\theta)^x \\ &= \sum_{x=0}^{\infty}\left[(1-\theta)^x-\theta x(1-\theta)^{x-1}\right] \\ &= \frac{1}{\theta}\sum_{x=0}^{\infty}\theta(1-\theta)^x-\frac{1}{1-\theta}\sum_{x=0}^{\infty}x\theta(1-\theta)^x \\ &= \frac{1}{\theta}-\frac{1}{1-\theta}E(X)=0.\end{aligned}$$

That is $E(X)=\frac{1-\theta}{\theta}$.

Justifying the interchange

Let $h(\theta, x) = \theta(1-\theta)^x$. Then $\frac{\partial}{\partial\theta} h(\theta, x) = (1-\theta)^x - \theta x(1-\theta)^{x-1}$, and verify that $\sum_{x=0}^{\infty} \frac{\partial}{\partial\theta} h(\theta, x)$ converges uniformly. Calculate $S_n(\theta)$ by

$$
\begin{aligned}
\lim_{n\to\infty} S_n(\theta) &= \lim_{n\to\infty} \left(\sum_{x=0}^{n} \left[(1-\theta)^x - \theta x (1-\theta)^{x-1} \right] \right) \\
&= \lim_{n\to\infty} \left(\frac{1-(1-\theta)^{n+1}}{\theta} - \theta \sum_{x=0}^{n} \frac{\partial}{\partial\theta} (1-\theta)^x \right) \\
&= \lim_{n\to\infty} \left(\frac{1-(1-\theta)^{n+1}}{\theta} - \theta \frac{\partial}{\partial\theta} \sum_{x=0}^{n} (1-\theta)^x \right) \\
&= \lim_{n\to\infty} \left(\frac{1-(1-\theta)^{n+1}}{\theta} - \theta \frac{\partial}{\partial\theta} \frac{1-(1-\theta)^{n+1}}{\theta} \right) \\
&= \lim_{n\to\infty} (n+1)(1-\theta)^n = 0.
\end{aligned}
$$

Take-aways

Conclusions

- Transformation
  - Functions of a r.v.
  - Monotone Transformations
- Expectation
  - Properties of Expectations
  - Moment
  - Moment Generating Functions
- Differentiating Under an Integral Sign