Proability and Statistics

MA20205

Bibhas Adhikari

Autumn 2022-23, IIT Kharagpur

Lecture 6
August 30, 2022

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\frac{d^2}{dt^2}M(t) = E\left(X^2e^{tX}\right)$

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\frac{d^2}{dt^2}M(t) = E\left(X^2e^{tX}\right)$ Similarly

(3) $\frac{d^n}{dt^n}M(t) = E\left(X^ne^{tX}\right)$

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\frac{d^2}{dt^2}M(t) = E\left(X^2e^{tX}\right)$ Similarly

(3) $\frac{d^n}{dt^n}M(t) = E\left(X^ne^{tX}\right)$

Setting $t = 0$,

$$\frac{d^n}{dt^n}M(t)\mid_{t=0} = E\left(X^ne^{tX}\right)\mid_{t=0} = E(X^n)$$

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\frac{d^2}{dt^2}M(t) = E\left(X^2e^{tX}\right)$ Similarly

(3) $\frac{d^n}{dt^n}M(t) = E\left(X^ne^{tX}\right)$

Setting $t = 0$,

$$\frac{d^n}{dt^n}M(t)\mid_{t=0} = E\left(X^ne^{tX}\right)\mid_{t=0} = E(X^n)$$

Theorem If $M(t) = a_0 + a_1t + a_2t^2 + \ldots + a_nt^n + \ldots$ is the Taylor expansion of $M(t)$ then $E(X^n) = n!a_n$ for all n.

Key parameters for analyzing random variables

Observation

(1) $\frac{d}{dt}M(t) = \frac{d}{dt}E(e^{tx}) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\frac{d^2}{dt^2}M(t) = E\left(X^2e^{tX}\right)$ Similarly

(3) $\frac{d^n}{dt^n}M(t) = E\left(X^ne^{tX}\right)$

Setting $t = 0$,

$$\frac{d^n}{dt^n}M(t) \mid_{t=0} = E\left(X^ne^{tX}\right) \mid_{t=0} = E(X^n)$$

Theorem If $M(t) = a_0 + a_1t + a_2t^2 + \ldots + a_nt^n + \ldots$ is the Taylor expansion of $M(t)$ then $E(X^n) = n!a_n$ for all n.
Proof is obvious.

Key parameters for analyzing random variables

Observation

(1) $\dfrac{d}{dt}M(t) = \dfrac{d}{dt}E(e^{tx}) = E\left(\dfrac{d}{dt}e^{tX}\right) = E\left(Xe^{tX}\right)$

(2) $\dfrac{d^2}{dt^2}M(t) = E\left(X^2 e^{tX}\right)$ Similarly

(3) $\dfrac{d^n}{dt^n}M(t) = E\left(X^n e^{tX}\right)$

Setting $t = 0$,

$$\frac{d^n}{dt^n}M(t) \mid_{t=0} = E\left(X^n e^{tX}\right) \mid_{t=0} = E(X^n)$$

Theorem If $M(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n + \ldots$ is the Taylor expansion of $M(t)$ then $E(X^n) = n! a_n$ for all n.

Proof is obvious.

Problems...

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at} M_X(t)$

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at} M_X(t)$

$$M_{X+a}(t) = E\left(e^{t(X+a)}\right) = E(e^{tX}e^{ta}) = e^{ta}E(e^{tX}) = e^{ta}M_X(t)$$

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at} M_X(t)$

$$M_{X+a}(t) = E\left(e^{t(X+a)}\right) = E(e^{tX}e^{ta}) = e^{ta}E(e^{tX}) = e^{ta}M_X(t)$$

(b) $M_{bX}(t) = M_X(bt)$

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at} M_X(t)$

$$M_{X+a}(t) = E\left(e^{t(X+a)}\right) = E(e^{tX}e^{ta}) = e^{ta}E(e^{tX}) = e^{ta}M_X(t)$$

(b) $M_{bX}(t) = M_X(bt)$

$$M_{bX}(t) = E^{(tb)X} = M_X(tb)$$

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at} M_X(t)$

$$M_{X+a}(t) = E\left(e^{t(X+a)}\right) = E(e^{tX}e^{ta}) = e^{ta}E(e^{tX}) = e^{ta}M_X(t)$$

(b) $M_{bX}(t) = M_X(bt)$

$$M_{bX}(t) = E^{(tb)X} = M_X(tb)$$

(c) $M_{\frac{X+a}{b}} = e^{\frac{a}{b}t} M_X\left(\frac{t}{b}\right)$

Key parameters for analyzing random variables

Theorem. Let $M_X(t)$ denote the mgf of a random variable X. Suppose $a, b \in \mathbb{R}$. Then

(a) $M_{X+a} = e^{at}M_X(t)$

$$M_{X+a}(t) = E\left(e^{t(X+a)}\right) = E(e^{tX}e^{ta}) = e^{ta}E(e^{tX}) = e^{ta}M_X(t)$$

(b) $M_{bX}(t) = M_X(bt)$

$$M_{bX}(t) = E^{(tb)X} = M_X(tb)$$

(c) $M_{\frac{X+a}{b}} = e^{\frac{a}{b}t}M_X\left(\frac{t}{b}\right)$
Proof follows from applying the above

Key parameters for analyzing random variables

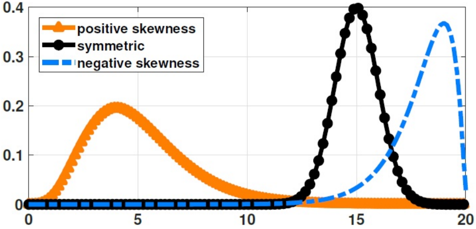
Skewness: measures the asymmetry of the distribution

$$\gamma = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$$

Key parameters for analyzing random variables

Skewness: measures the asymmetry of the distribution

$$\gamma = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$$



Key parameters for analyzing random variables

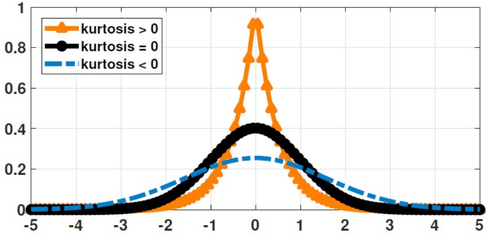
Kurtosis: measures how heavey-tailed the distribution is

$$\kappa = E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right)$$

Key parameters for analyzing random variables

Kurtosis: measures how heavey-tailed the distribution is

$$\kappa = E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right)$$



Positive and negative kurtosis is defined based on the kurtosis of Gaussian random variable

Special discrete distributions

Bernoulli random variable Let X be a random variable with range space $R_X = \{0, 1\}$. Then we say X to have Bernoulli distribution if the pmf of X is

$$f(0) = 1 - p \text{ and } f(1) = p$$

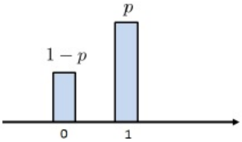
for some $0 < p < 1$.

Special discrete distributions

Bernoulli random variable Let X be a random variable with range space $R_X = \{0, 1\}$. Then we say X to have Bernoulli distribution if the pmf of X is

$$f(0) = 1 - p \text{ and } f(1) = p$$

for some $0 < p < 1$.

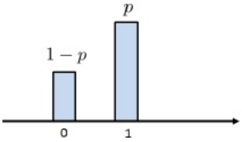


Special discrete distributions

Bernoulli random variable Let X be a random variable with range space $R_X = \{0, 1\}$. Then we say X to have Bernoulli distribution if the pmf of X is

$$f(0) = 1 - p \text{ and } f(1) = p$$

for some $0 < p < 1$.



Then we denote $X \sim \text{Bernoulli}(p)$

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

(b) $E(X^2) = p$

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

(b) $E(X^2) = p$

(c) $\text{Var}(X) = p(1 - p)$

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

(b) $E(X^2) = p$

(c) $\text{Var}(X) = p(1-p)$

(d) $M(t) = (1-p) + pe^t$

Application Social network modelling, the existence of a link can be modelled as the Bernoulli random variable

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

(b) $E(X^2) = p$

(c) $\text{Var}(X) = p(1-p)$

(d) $M(t) = (1-p) + pe^t$

Application Social network modelling, the existence of a link can be modelled as the Bernoulli random variable

Question For what values of p the $\text{Var}(X)$ attains the maximum value?

Special discrete distributions

Observation The Bernoulli(p) models the phenomena of coin toss/true-false/yes-no etc.

Properties of Ber(p) :

(a) $E(X) = p$

(b) $E(X^2) = p$

(c) $\text{Var}(X) = p(1-p)$

(d) $M(t) = (1-p) + pe^t$

Application Social network modelling, the existence of a link can be modelled as the Bernoulli random variable

Question For what values of p the $\text{Var}(X)$ attains the maximum value?

Special discrete distributions

Binomial random variable A random variable X is said to have binomial distribution if the pmf of X is

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \ldots, n$$

where $0 < p < 1$ and n is the total number of states.

Special discrete distributions

Binomial random variable A random variable X is said to have binomial distribution if the pmf of X is

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \ldots, n$$

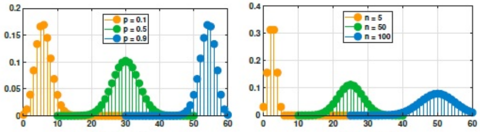
where $0 < p < 1$ and n is the total number of states.
Then we write $X \sim \text{Binomial}(n, p)$

Special discrete distributions

Binomial random variable A random variable X is said to have binomial distribution if the pmf of X is

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$$

where $0 < p < 1$ and n is the total number of states.
Then we write $X \sim \text{Binomial}(n, p)$
The pmf of Binomial(n, p) looks like:



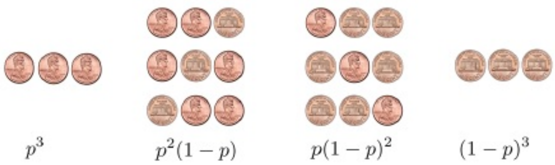
(a) $n = 60$ (b) $p = 0.5$

Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$

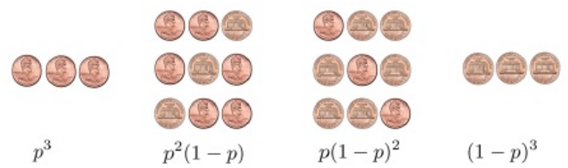
Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$



Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$

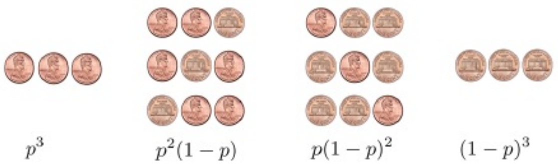


Properties of Bin(n, p)

(a) $E(X) = np$

Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$



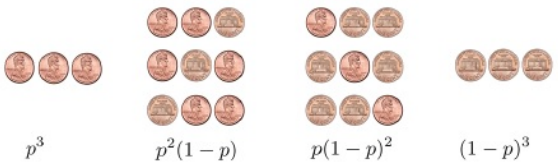
Properties of Bin(n, p)

(a) $E(X) = np$

(b) $E(X^2) = np(np + (1 - p))$

Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$



Properties of Bin(n, p)

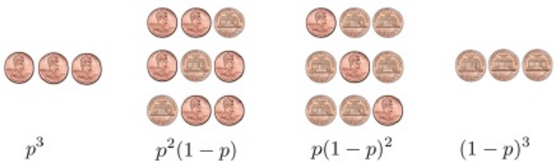
(a) $E(X) = np$

(b) $E(X^2) = np(np + (1 - p))$

(c) $\mathrm{Var}(X) = np(1 - p)$

Special discrete distributions

Observation Binomial(n, p) explains the phenomena of $X = k$ number of heads in n coin toss, when the probability of getting head is p whereas the tail is $1 - p$



Properties of Bin(n, p)

(a) $E(X) = np$

(b) $E(X^2) = np(np + (1 - p))$

(c) $\text{Var}(X) = np(1 - p)$

(d) $M(t) = [(1 - p) + pe^t]^n$