# Special random variables

June 8, 2023

# Table of contents

# Special Discrete RVs

- Understand the assumptions for some common discrete probability distributions
- Select an appropriate discrete probability distribution to calculate probabilities in specific applications
- Calculate probabilities, determine means and variances for some common discrete probability distributions

# Bernoulli RV

Discrete RV $X$ is called *Bernoulli RV* with parameter $p$ if its pmf is

$$P(X = 0) = 1 - p$$
$$P(X = 1) = p$$

Denote $X \sim \text{Ber}(p)$

# Mean and Variance of Bernouilli RV

$X \sim \text{Ber}(p)$

$$E(X) = p$$

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

# Use Bernoulli RV to

model generic probabilistic situations with just two outcomes:

- The state of a telephone at a given time that can be either free or busy.
- A person who can be either healthy or sick with a certain disease.

# Use Bernoulli RV to

construct more complicated RV by combining multiple Bernoulli RV

# Geometric RV

- toss a biased coin
- $P(Head) = p$, $P(Tail) = 1 - p$
- $X$: number of tosses until a head comes up for the first time
- pmf of $X$

$$p(k) = (1-p)^{k-1}p, \ k \geq 1$$

- $X$ is Geometric with parameters $p$, denoted by $X \sim Geo(p)$

Repeat independent Bernoulli trials until the first success

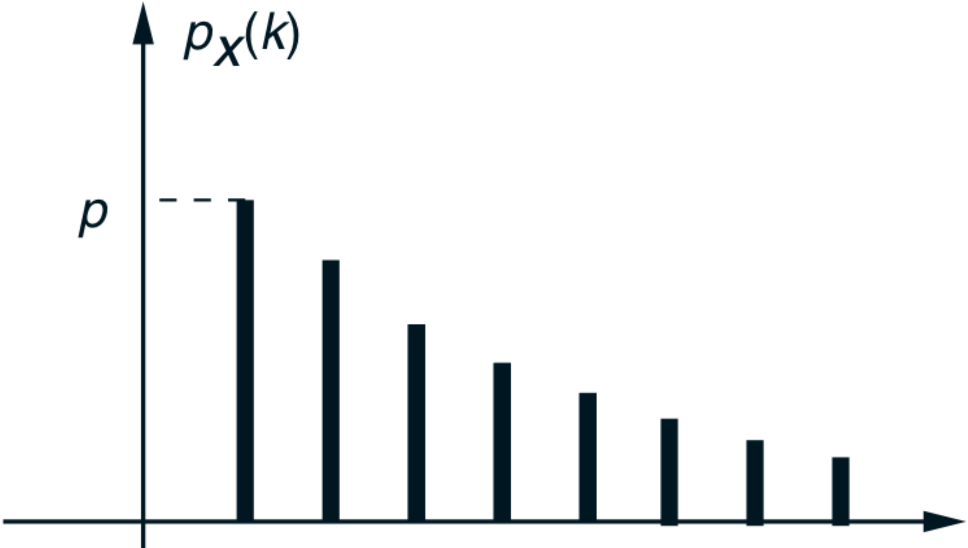
# Mean and Variance of Geometric RV

$X \sim Geo(p)$

$$E(X) = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1}{p^2}$$

# pmf of Geometric RV is decreasing



# Example - Digital channel

The chance that a bit transmitted through a digital transmission channel is received in error is .1. Also, assume that the transmission trials are independent. $X$ denote the number of bits transmitted until the first error. Determine $P(X = 5)$.

# Solution

- $X \sim Geo(.1)$
- $$P(X = 5) = (.9)^4(.1) \approx .066$$

# Binomial RV

- toss a biased coin $n$ time
- $P(Head) = p$, $P(Tail) = 1 - p$
- $X$: number of heads in the $n$-toss sequence
- pmf of $X$

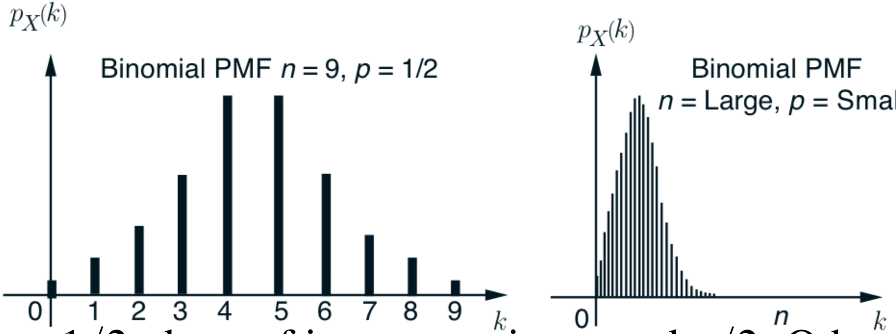
$$p(k) = \binom{n}{k} p^k (1-p)^{n-k},\ 0 \le k \le n$$

- $X$ is Binomial with parameters $(n, p)$, denoted by $X \sim \text{Bino}(n, p)$

# Meaning

- Counting the number of success in an experiment consisting of $n$ independent Bernoulli trials
- sum of $n$ independent and identical Bernouilli RV

# pmf of Binomial RV




If $p = 1/2$, the pmf is symmetric around $n/2$. Otherwise, the pmf is skewed towards 0 if $p < 1/2$, and towards $n$ if $p > 1/2$.

# Mean and Variance of Binomial RV

$X \sim \text{Bino}(n, p)$

$$E(X) = np$$

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

# Example - Digital channel

The chance that a bit transmitted through a digital transmission channel is received in error is .1. Also, assume that the transmission trials are independent. Let $X$ the number of bits in error in the next four bits transmitted. Determine $P(X = 2)$.

# Solution

- $X \sim \text{Bino}(4, .1)$
- $$P(X = 2) = \binom{4}{2}(.1)^2(.9)^2 \approx .0486$$

# Table of contents

## Normal distribution

- Use the table for the cumulative distribution function of a standard normal distribution to calculate probabilities
- Standardize normal random variables

Continuous RV $X$ is said to be normally distributed or Gaussian with parameter $\mu$ and $\sigma^2$ if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

for $-\infty < x < \infty$
Denote $X \sim \mathcal{N}(\mu, \sigma^2)$.

# Mean and variance of $\mathcal{N}(\mu, \sigma^2)$

$X \sim \mathcal{N}(\mu, \sigma^2)$

- $E(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$

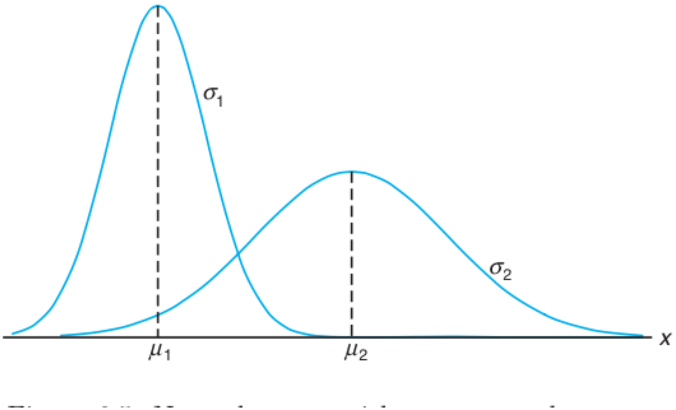



Figure: $\mu_1 < \mu_2$, $\sigma_1 < \sigma_2$

- Normal distibution is the most widely used distribution
- Many random phenomena obey a normal distribution
- Ex: the height and weight of a person, accuracy of shots from a gun... Link

# Good approximation

- to approximate Binomial $(n, p)$ when $n$ is large
- *limiting distribution* of sample mean ...**broad base** for statistic inference (estimation and hypothesis testing), analysis of variance

# Standard normal distribution

- $Z \sim \mathcal{N}(0,1)$ is standard normal distribution
- pdf

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- cdf

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

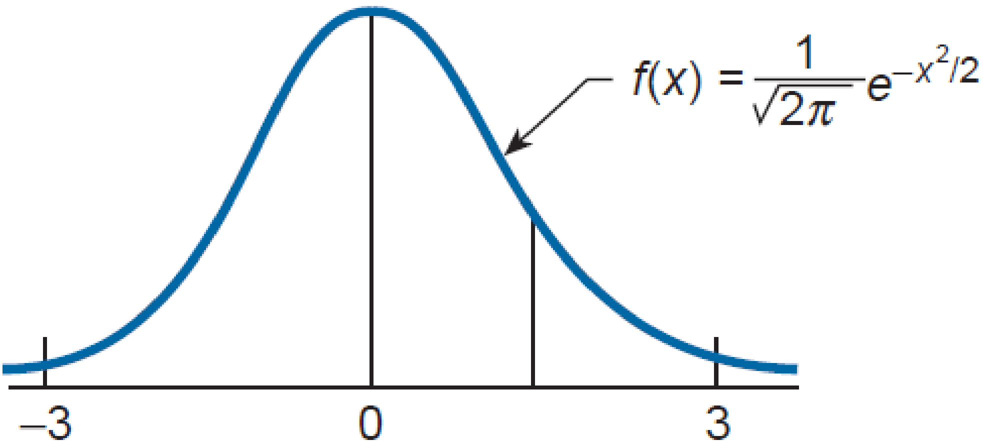



Figure: Pdf of $Z$

# Compute probability of standard normal distribution

- Calculator
- Look up values in Normal Probability Table

# Standard normal probability table (cdf)



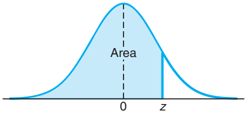



**Table A.3** Areas under the Normal Curve

| $z$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| **−3.4** | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |
| **−3.3** | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| **−3.2** | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| **−3.1** | 0.0010 | 0.0009 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| **−3.0** | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| **−2.9** | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |

# Example

$$P(Z \leq -2.54)$$

# Solution 1 - Calculator

1.

$$P(Z \leq -2.54) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-2.54} e^{-\frac{x^2}{2}} dx$$

$$= \lim_{a \to -\infty} \frac{1}{\sqrt{2\pi}} \int_{a}^{-2.54} e^{-\frac{x^2}{2}} dx$$

Substitute $a$ by $-10$, $-30$, $-50$ ... and find the limit

# Solution 2 - Look up the table value of normal probability

1. look up $-2.5$ in the first column
2. look up .04 in the first row
3. Intersection of the corresponding row and column

$$P(Z < -2.54) = .0055$$

# Property

$$P(Z \leq z) = \phi(z)$$
$$P(Z > z) = 1 - \phi(z)$$
$$P(a < Z < b) = \phi(b) - \phi(a)$$
$$P(-a < Z < a) = 2\phi(a) - 1 \text{ for } a > 0$$

Find

1. $P(Z > 2.33)$
2. $P(-1.65 < Z < 1.65)$
3. $z$ such that $P(Z > z) = .95$

A bit 1 is sent from location A to location B. The value recieved at B is $R = 1 + Z$ where $Z$ is the channel noise disturbance. When the message is received at location B, the receiver decodes it according to the following rule: If $R \geq 0.5$ then "1" is concluded to be sent. If $R < 0.5$ then "0" is concluded. What is the probability that the decode is incorrect?

# Normality is Preserved by Linear Transformations

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then
$Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

# Standardize a normal distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma})$$
$$= P(Z \leq \frac{x-\mu}{\sigma}) = \phi(\frac{x-\mu}{\sigma})$$

and

$$P(a \leq X \leq b) = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$$

# Example

The annual snowfall at a particular geographic location is modeled as a normal random variable with a mean of $\mu = 60$ inches, and a standard deviation of $\sigma = 20$. What is the probability that this year's snowfall will be at least 80 inches?

# Solution

- Snowfall $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 60$, $\sigma = 20$
- 

$$P(X \geq 80) = 1 - P(X < 80)$$
$$= 1 - \phi(\frac{80 - \mu}{\sigma}) = 1 - .8413 = .1687$$

# Practice

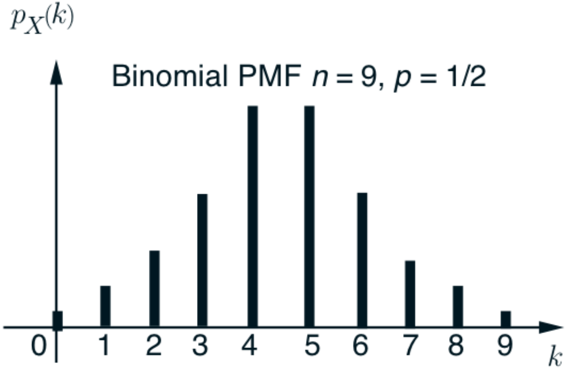
The power $W$ dissipated in a resistor is proportional to the square of the voltage $V$

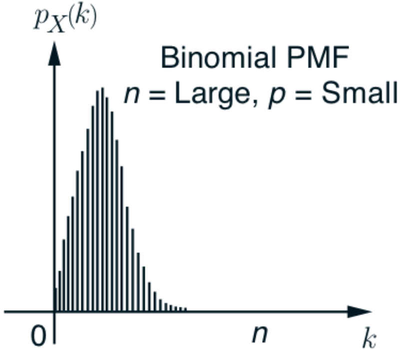
$$W = 3V^2$$

Suppose $V \sim \mathcal{N}(6, 1)$. Compute $P(W > 120)$

# Table of contents

# Binomial and normal



# Binomial approximation

- Suppose $Y \sim \text{Bino}(n, p)$ where $n$ is large and $np$ is not too small
- $Y$ can be approximated by
  $X \sim \mathcal{N}(\underbrace{np}_{E(Y)}, \underbrace{np(1-p)}_{Var(Y)})$
- $Y$ is discrete, $X$ is continuous
- so we have to "fill the gap"

# "Fill the gap" - midpoint rule

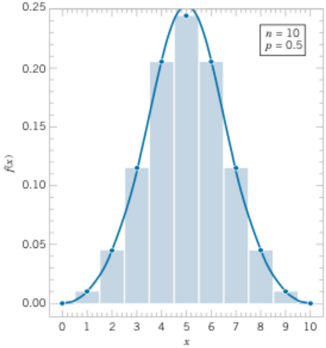



**Figure 4-19** Normal approximation to the binomial distribution.

$$P(Y = i) \approx P(i - \frac{1}{2} < X < i + \frac{1}{2})$$

# Continuity correction

To approximate a binomial probability of $X \hookrightarrow Bin(n, p)$ with a normal distribution, a **continuity correction** is applied as follows:

$$P(X \leq x) = P(X \leq x + 0.5) \approx P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

*The approximation is good for* $np > 5$ *and* $n(1-p) > 5$.

Toss a fair coin 40 times. $Y$ is the number of heads. Calculate $P(Y = 20)$ using normal approximation and direct computation.

# Solution

Approximate $Y$ by $X \sim \mathcal{N}(20, 10)$

$$
\begin{aligned}
P(Y = 20) &\approx P(19.5 < X < 20.5) \\
&= P\left(\frac{19.5 - 20}{\sqrt{10}} < Z < \frac{20.5 - 20}{\sqrt{10}}\right) \\
&= P(Z < .16) - P(Z < -.16) \\
&= .1272
\end{aligned}
$$

# Exact value

$$P(Y = 20) = \binom{40}{20}(.5)^{40} = .1254$$

# Example

The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

# Solution

$X$: number of attending students

$X \sim \text{Bino}(450, .3) \approx \mathcal{N}(135, 94.5)$

$$
\begin{aligned}
P(X > 150) &= P(X \geq 150.5) \\
&\approx P(Z \geq \frac{150.5 - 135}{\sqrt{94.5}}) \\
&= 1 - P(Z < 1.59) \\
&= .0559 \approx 5.6\%
\end{aligned}
$$

# Normal Approximation to the Poisson Distribution

If $X$ is a Poisson random variable with $E(X) = \lambda$ and $Var(X) = \lambda$ then for $\lambda > 5$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable.

# Continuity correction

$X \sim Pois(\lambda)$ then for $\lambda > 5$,

$$P(X \leq x) = P(X \leq x + 0.5)$$
$$\approx P\left(Z \leq \frac{x + 0.5 - \lambda}{\sqrt{\lambda}}\right)$$

# Example

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

# Solution

Let $X \sim Pois(1000)$ then we need to compute

$$P(X \leq 950) = \sum_{k=0}^{950} e^{-1000} \frac{1000^k}{k!}$$

which is difficult to compute but can be approximated by

$$\begin{aligned} P(X \leq 950) &= P(X \leq 950.5) \\ &\approx P\left(Z \leq \frac{950.5 - 1000}{\sqrt{1000}}\right) \\ &= P(Z \leq -1.57) = 0.058 \end{aligned}$$