

# Probability & Statistics for Data Science

David Li

# Random Variable and Distribution

- A *random variable*  $X$  is a numerical outcome of a random experiment
- The *distribution* of a random variable is the collection of possible outcomes along with their probabilities:
  - Discrete case:  $\Pr(X = x) = p_{\theta}(x)$
  - Continuous case:  $\Pr(a \leq X \leq b) = \int_a^b p_{\theta}(x)dx$
- Distribution simulation
  - <https://www.youtube.com/watch?v=6YDHBfVivIs>

# Random Variables Distributions

- Cumulative Probability Distribution (CDF):

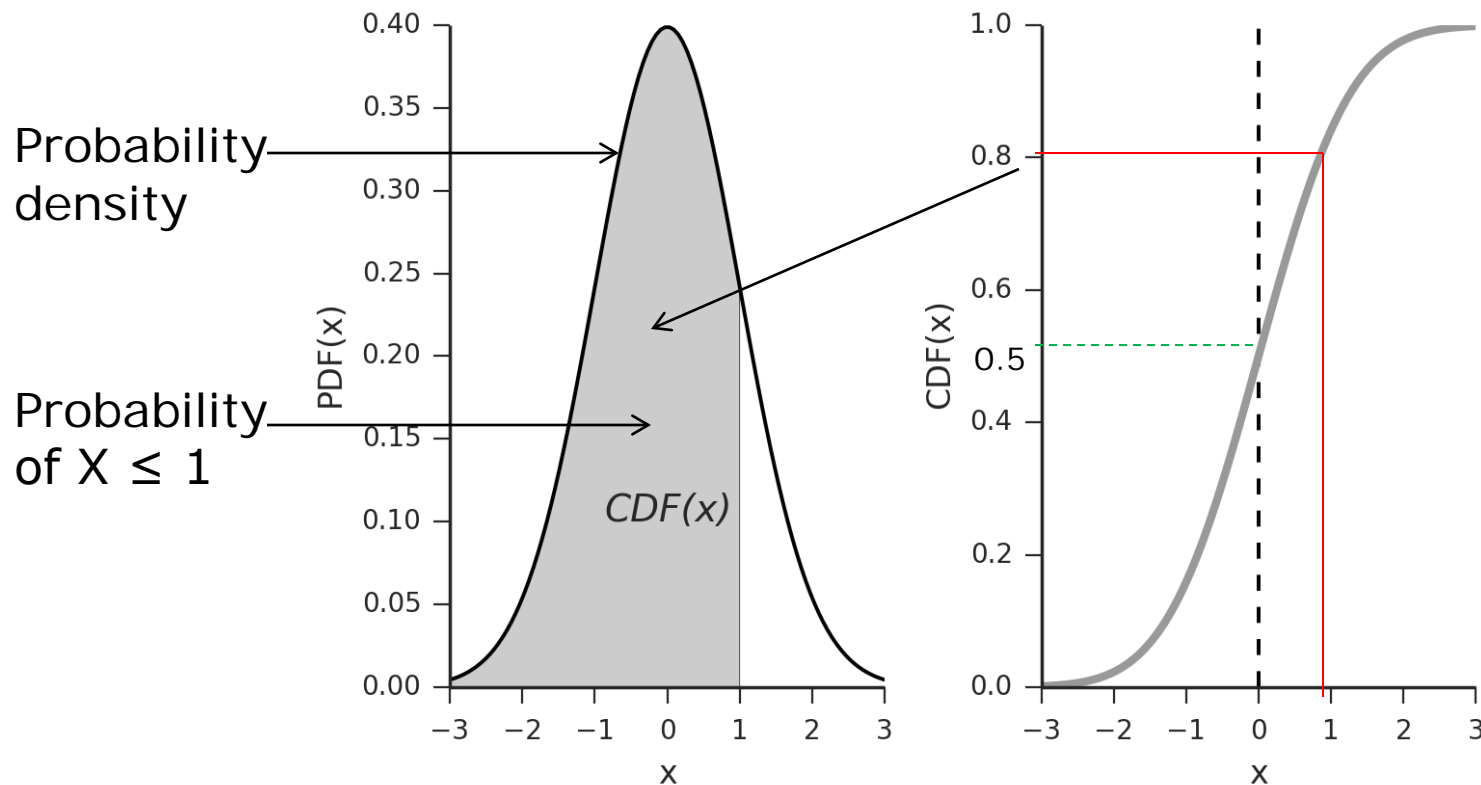
$$F_X(x) = P(X \leq x)$$

- Probability Density Function (PDF):

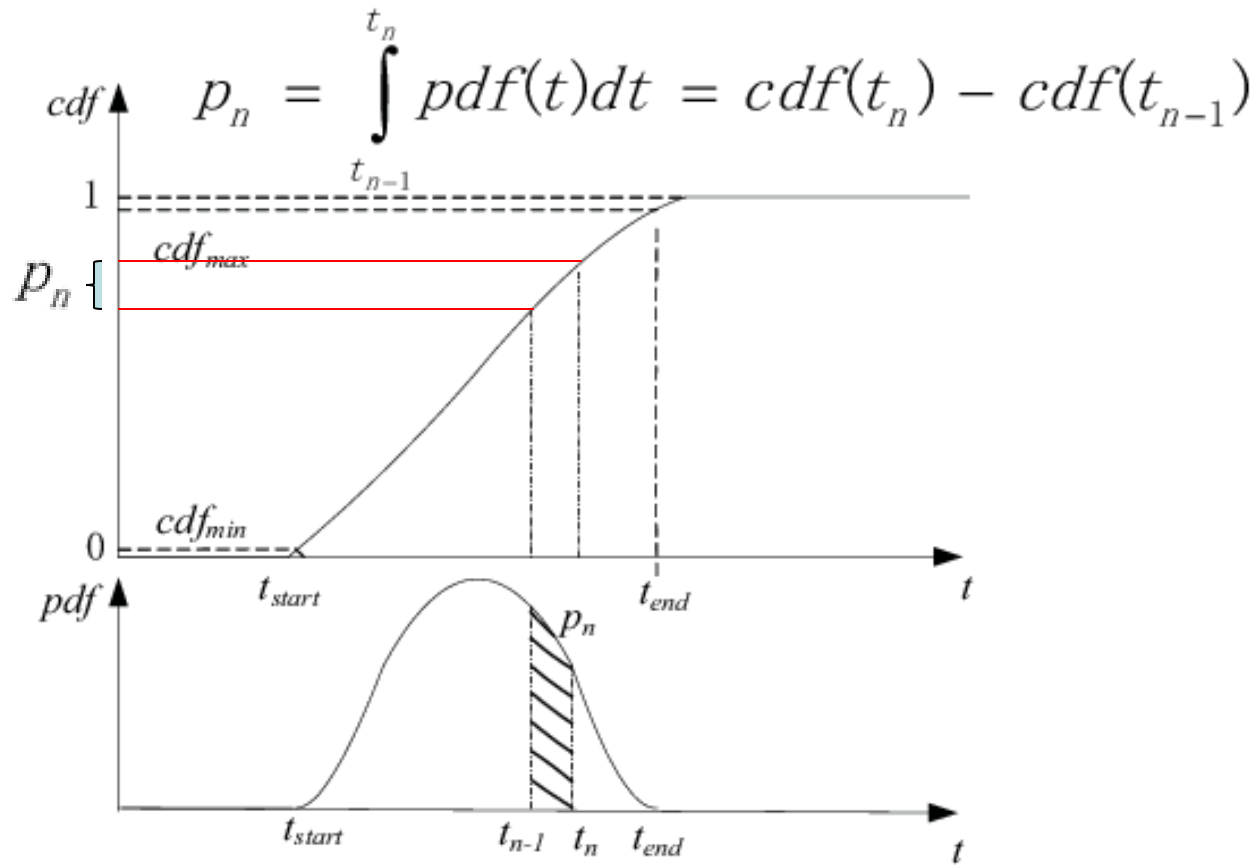
$$p_X(x) = \frac{dF_X(x)}{dx}$$

$$\int_{-\infty}^{\infty} p_X(x) dx = 1.0$$

# Random Variables Distributions



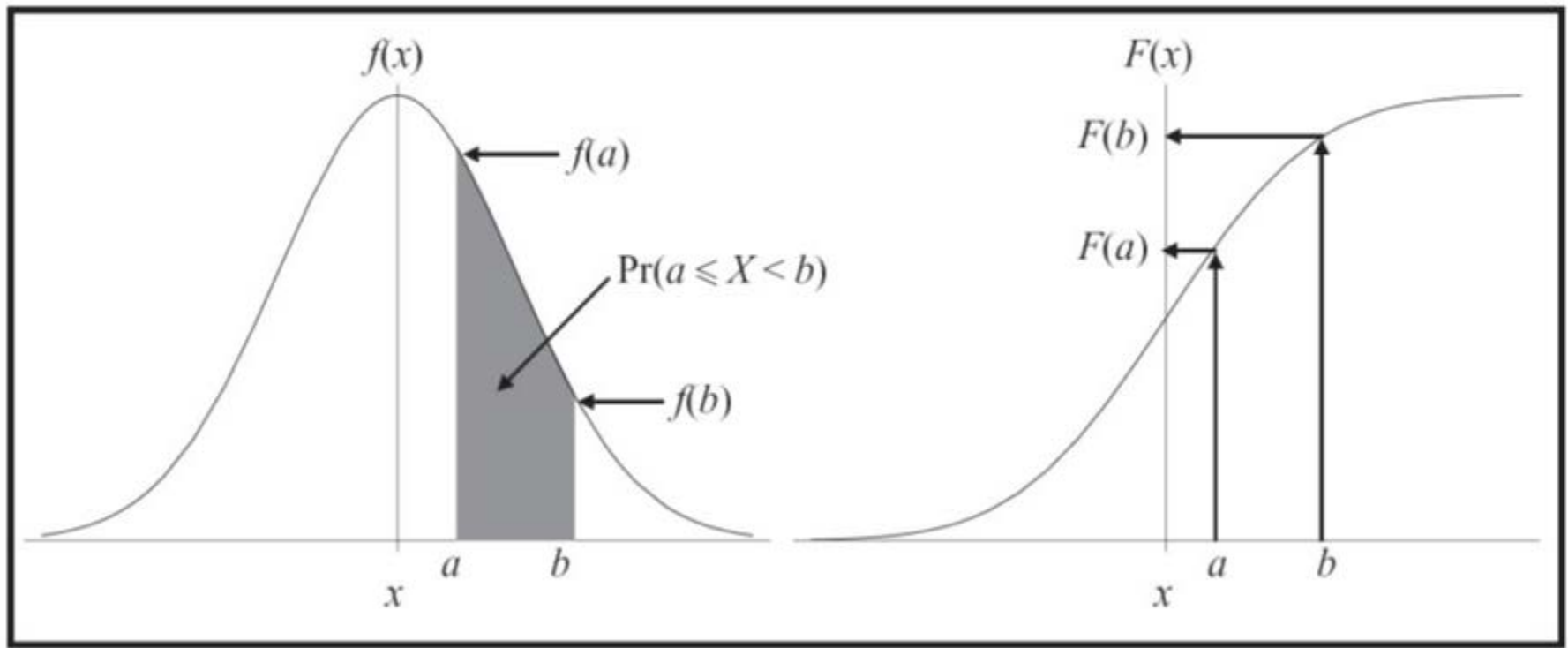
# Random Variables Distributions



# Random Variables Distributions

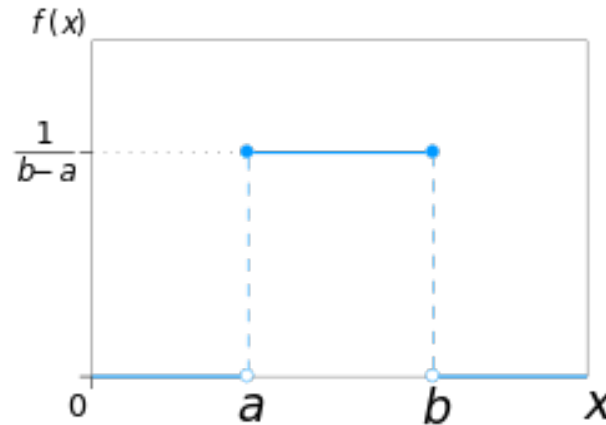
Probability density function

Cumulative distribution function



# A frequently asked question:

- ❑ Can a probability density function (pdf) be greater than 1?
- ❑ Let's take an example of the easiest PDF — the uniform distribution defined on the domain  $[a, b]$ . The PDF of the uniform distribution is  $1/(b-a)$  within the domain, which is constantly 2 throughout when  $b=1$  and  $a=0.5$ , and 0 elsewhere



- ❑ The total probability is the total area under the graph  $f(x)$ , which is  $2 * 0.5 = 1$ .

# Random Variable: Example

- Let  $S$  be the set of all sequences of three rolls of a dice. Let  $X$  be the sum of the number of dots on the three rolls.
- What are the possible values for  $X$ ?
- $\Pr(X = 5) = ?$ ,  $\Pr(X = 10) = ?$

- **Solution**

$$5 = 1 + 1 + 3 = 1 + 2 + 2$$

$$(3+3)/(6*6*6) = 1/36$$

$$10 = 1 + 3 + 6 = 2 + 2 + 6 = 2 + 3 + 5 = 2 + 4 + 4$$

$$(6+3+6+3)/(6*6*6) = 18/(6*6*6) = 1/12$$



# Expectation

- Definition: the expectation of a random variable is

$$E[X] = \sum_x x \Pr(X = x) \quad \text{discrete case}$$

What is "average"?

$$E[X] = \int_{-\infty}^{\infty} x p_{\theta}(x) dx \quad \text{continuous case}$$

- Properties

- Summation: For any  $n \geq 1$ ,  $X_1, X_2, \dots, X_n$  may be dependent and any constants  $k_1, \dots, k_n$

$$E\left[\sum_{i=1}^n k_i X_i\right] = \sum_{i=1}^n k_i E(X_i)$$

- Product: If  $X_1, X_2, \dots, X_n$  are independent

$$E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E(X_i)$$

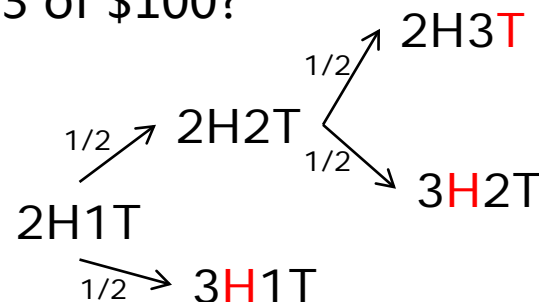
# Expectation

- Division of the stakes problem

Henry and Tony play a game. They toss a fair coin, if get a Head, Henry wins; Tail, Tony wins. They contribute equally to a prize pot of \$100, and agree in advance that the first player who has won 3 rounds will collect the entire prize. However, the game is interrupted for some reason after 3 rounds. They got 2 H and 1 T. How should they divide the pot fairly?

- It seems unfair to divide the pot equally Since Henry has won 2 out of 3 rounds. Then, how about Henry gets 2/3 of \$100?
- Other thoughts?

X	0	100
P	0.25	0.75



X is what Henry will win if the game not interrupted

# Expectation: Example

- Let  $S$  be the set of all sequence of three rolls of a dice. Let  $X$  be the sum of the number of dots on the **three** rolls.
- What is  $E(X)$ ?
  - $E[X1+X2+X3] = E[X1]+E[X2]+E[X3]=3.5*3=10.5$
- Let  $S$  be the set of all sequence of three rolls of a dice. Let  $X$  be the product of the number of dots on the **three** rolls.
- What is  $E(X)$ ?
  - $E[X1*X2*X3] = E[X1]*E[X2]*E[X3]=3.5^3$

# Variance

- Informally, variance measures how far a set of (random) numbers are spread out from their average value.
- Mathematically, the variance of a random variable  $X$  is the expectation of  $(X-E[X])^2$  :

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2\end{aligned}$$

- Properties
  - For any constant  $C$ ,  $\text{Var}(CX) = C^2\text{Var}(X)$
  - If  $X_1, X_2, \dots, X_n$  are independent  
 $\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$

# Standard Deviation

- Definition: the **square root** of the **Variance**:

The "**Population** Standard Deviation": 
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

The "**Sample** Standard Deviation": 
$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Looks complicated, but the important change is to divide by **N-1** (instead of **N**) when calculating a Sample Variance.

# Covariance and Correlation

$$COV(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

x = the independent variable

y = the dependent variable

n = number of data points in the sample

$\bar{x}$  = the mean of the independent variable x

$\bar{y}$  = the mean of the dependent variable y

From Wikipedia:

Covariance is a measure of the joint variability of two random variables. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, the covariance is positive

$$r_{(x,y)} = \frac{COV(x,y)}{s_x s_y}$$

r(x,y) = correlation of the variables x and y

COV(x, y) = covariance of the variables x and y

s<sub>x</sub> = sample standard deviation of the random variable x

s<sub>y</sub> = sample standard deviation of the random variable y

Normalize covariance to the range [-1, 1] to make it comparable

# Standard Deviation of two correlated variables

$$\text{Cor}(R_i, R_j) = \frac{\text{Cov}(R_i, R_j)}{\sigma_i \sigma_j}$$

$$\sigma_p = \sqrt{w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \rho_{1,2} \sigma_1 \sigma_2}$$

$$\sigma_p = \sqrt{w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \text{Cov}_{1,2}}$$

# Example: stock portfolio

- Suppose you want to invest your wealth in the stocks of ABC, Inc. and XYZ, Inc. The expected returns and standard deviations of the returns for the two corporations are

$$E(r_{ABC}) = 20\% \quad \sigma_{ABC} = .26$$

$$E(r_{XYZ}) = 10\% \quad \sigma_{XYZ} = .16$$

- Now assume that the covariance of the returns from the two companies is zero
- What are the expected return and standard deviation of your portfolio if you invested 50 percent in ABC, Inc. and 50 percent in XYZ, Inc.



# Example: stock portfolio answers

$$\begin{aligned} E(r_{\text{Port}}) &= .5(20\%) + .5(10\%) \\ &= 15\% \end{aligned}$$

$$\begin{aligned} \rho_{i,j} &= \frac{\text{COV}(r_i, r_j)}{\sigma_i \sigma_j} \\ &= \frac{0}{.26 \times .16} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \sigma_{\text{Port}} &= \left( \sigma_i^2 W_i^2 + \sigma_j^2 W_j^2 + 2r_{i,j} \sigma_i \sigma_j W_i W_j \right)^{\frac{1}{2}} \\ &= \left[ (.26)^2 (.5)^2 + (.16)^2 (.5)^2 \right]^{\frac{1}{2}} \\ &\quad + 2(0)(.26)(.16)(.5)(.5) \\ &= .1526 \end{aligned}$$

# Bernoulli Distribution

- The outcome of an experiment can either be success (i.e., 1) and failure (i.e., 0).
- $\Pr(X=1) = p$ ,  $\Pr(X=0) = 1-p$ , or

$$p_{\theta}(x) = p^x (1-p)^{1-x}$$

- $E[X] = p$ ,  $\text{Var}(X) = p(1-p)$
- Bernoulli random variable can take either 0 or 1 using certain probability as a parameter.
- To generate 10000 Bernoulli random numbers with success probability  $p=1/4$ , we will use `sample()`

# Bernoulli Distribution

```
n = 10; p = 1/4;  
sample(0:1, size=n, replace=TRUE, prob=c(1-p, p))
```

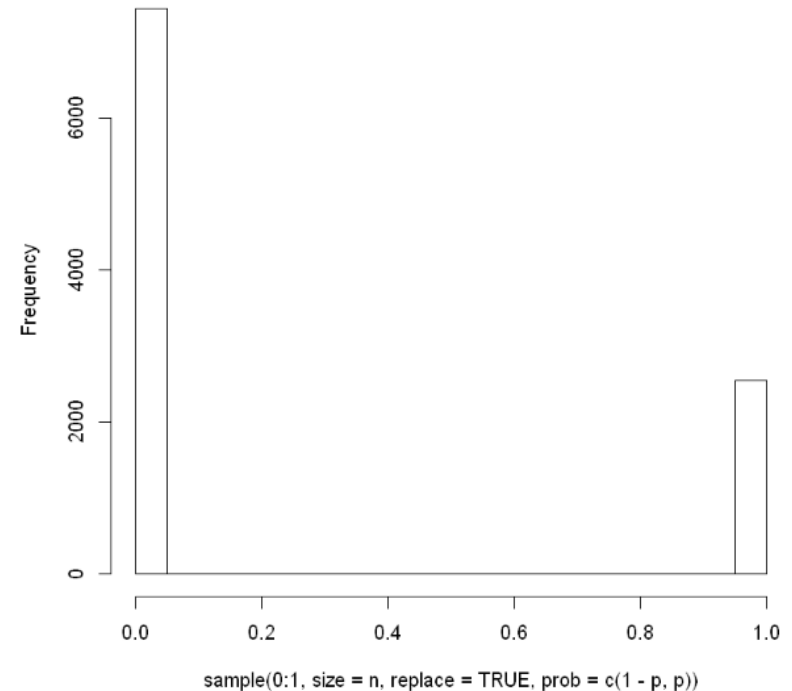
```
0 0 0 0 1 1 0 0 0 0
```

```
n = 10000; p = 1/4;  
hist(sample(0:1, size=n, replace=TRUE, prob=c(1-p, p)))
```

```
n = 10000; p = 1/4;  
table(sample(0:1, size=n, replace=TRUE, prob=c(1-p, p)))
```

```
0    1  
7496 2504
```

Histogram of sample(0:1, size = n, replace = TRUE, prob = c(1 - p, p))



# Binomial Distribution

- n draws of a Bernoulli distribution
  - $X_i \sim \text{Bernoulli}(p)$ ,  $X = \sum_{i=1}^n X_i$ ,  $X \sim \text{Bin}(p, n)$
- Random variable X stands for the number of times that experiments are successful.

$$\Pr(X = x) = p_{\theta}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

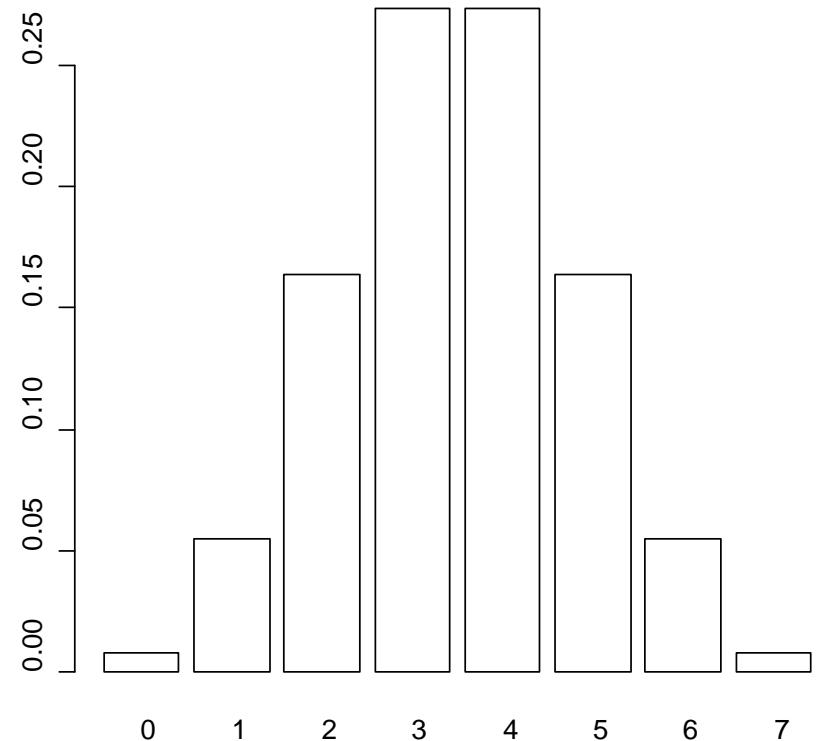
- n = the number of trials
  - x = the number of successes
  - p = the probability of success
- $E[X] = np$ ,  $\text{Var}(X) = np(1-p)$ 
  - Intuitively binomial just repeats bernoulli n times

# the binomial distribution in R

- `dbinom(x, size, prob)`  
where “d” means density
- Try 7 times, equally likely succeed or fail.  
The prob of succeeding 3 times in the 7 tries.

```
> dbinom(3,7,0.5)  
[1] 0.2734375
```

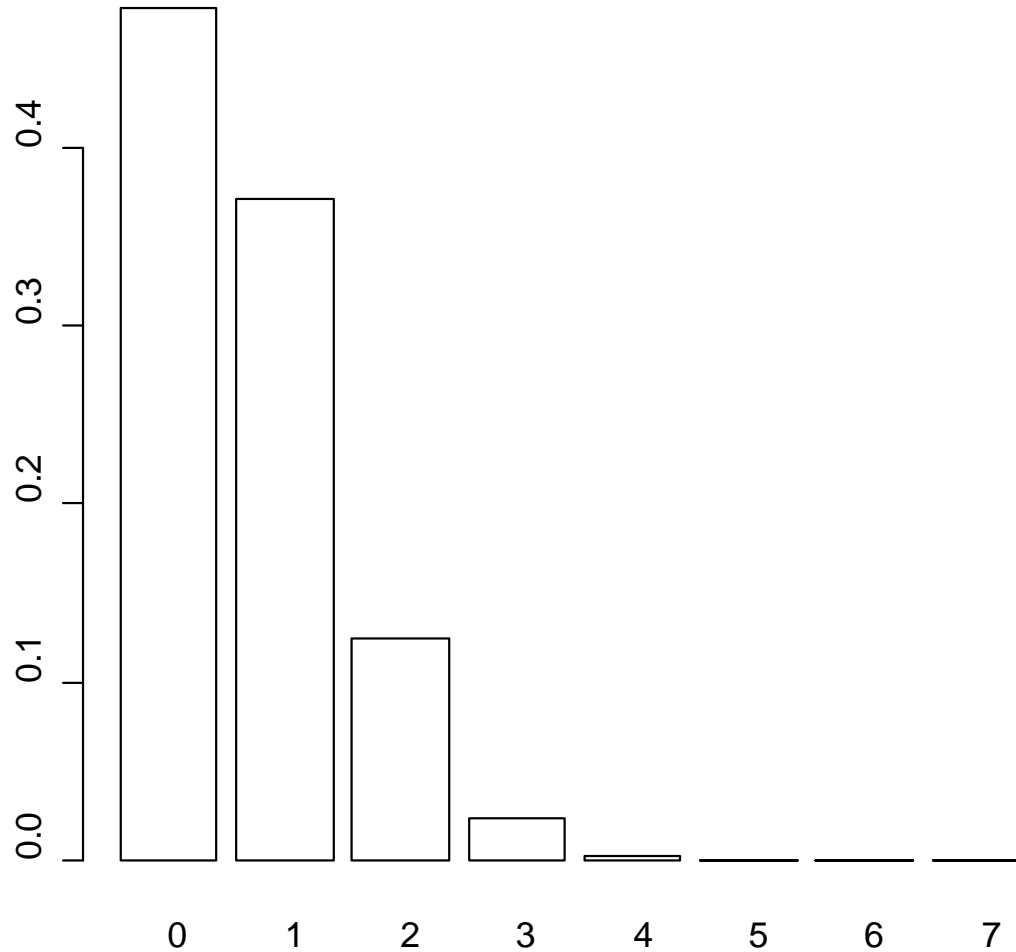
```
> barplot(dbinom(0:7,7,0.5),  
names.arg=0:7)
```



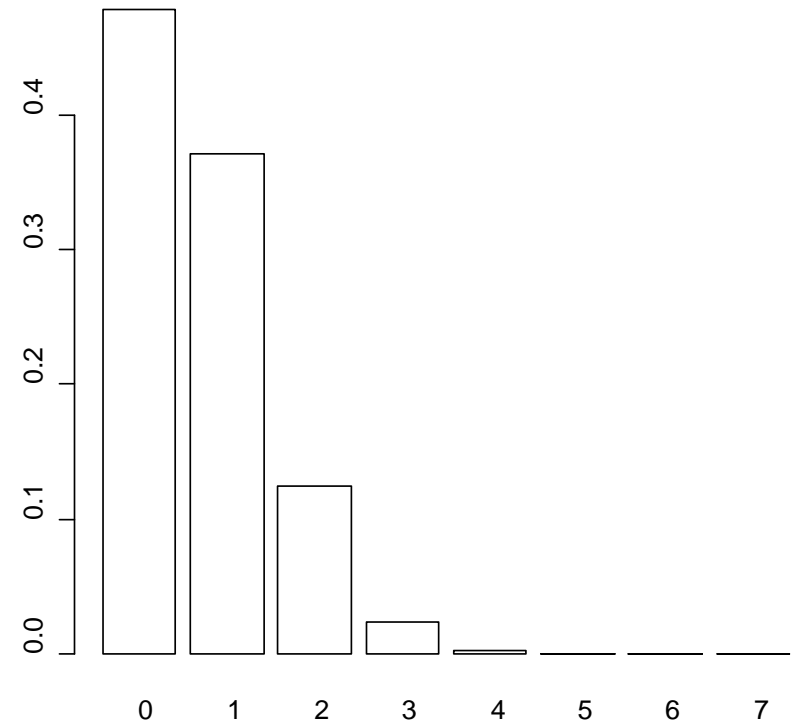
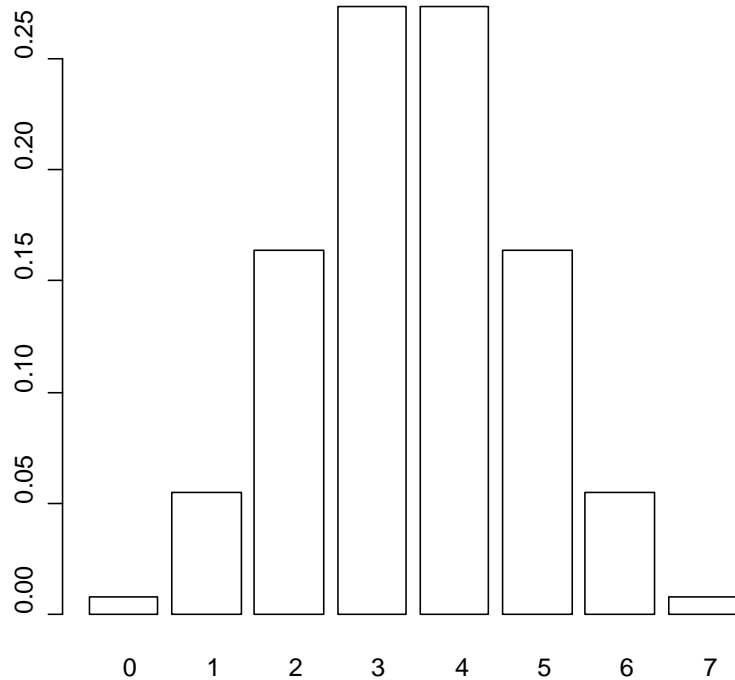
Why is the plot symmetric?

# what if $p \neq 0.5$ ?

- `> barplot(dbinom(0:7,7,0.1),names.arg=0:7)`



# Which distribution has greater variance?

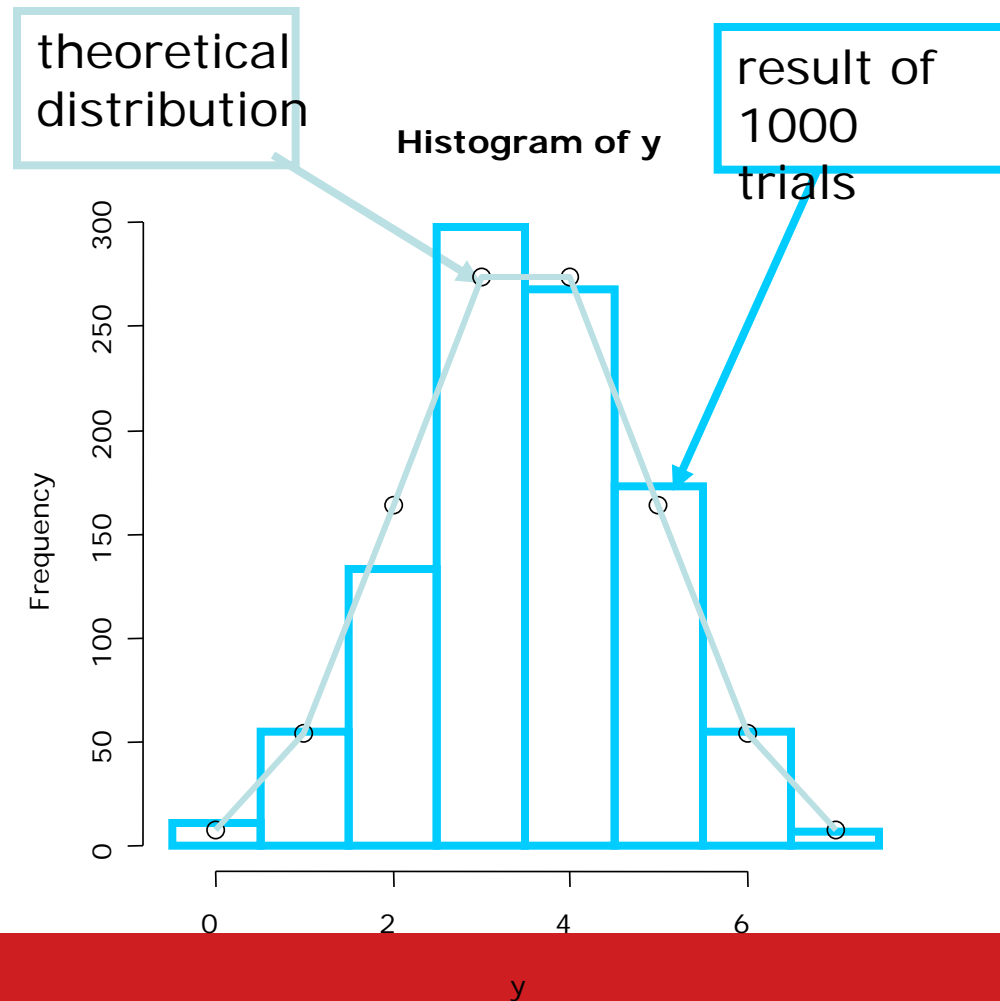


$$\text{var} = n \cdot p \cdot (1-p) \quad p = 0.5 \\ \text{var} = 7 \cdot 0.5 \cdot 0.5 = 7 \cdot 0.25$$

$$\text{var} = n \cdot p \cdot (1-p) \quad p = 0.1 \\ \text{var} = 7 \cdot 0.1 \cdot 0.9 = 7 \cdot 0.09$$

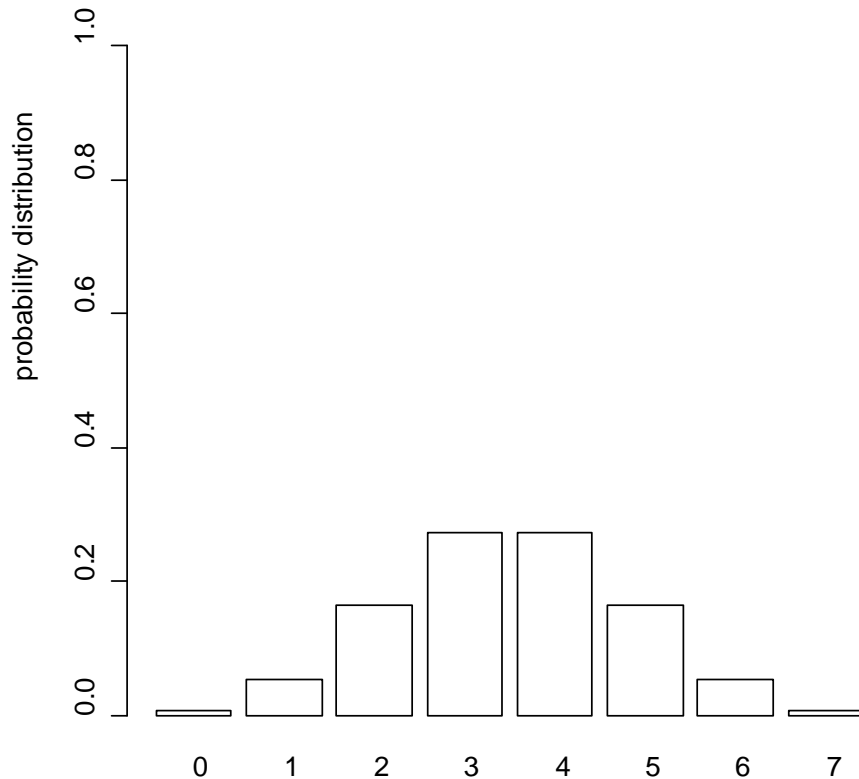
# briefly comparing an experiment to a distribution

```
nExpr = 1000
tosses = 7; y=rep(0,nExpr);
for (i in 1:nExpr) {
  x = sample(c("H","T"),
    tosses, replace = T)
  y[i] = sum(x=="H")
}
hist(y,breaks=-0.5:7.5)
lines(0:7,dbinom(0:7,7,0.5)*nExpr)
points(0:7,dbinom(0:7,7,0.5)*nExpr)
```

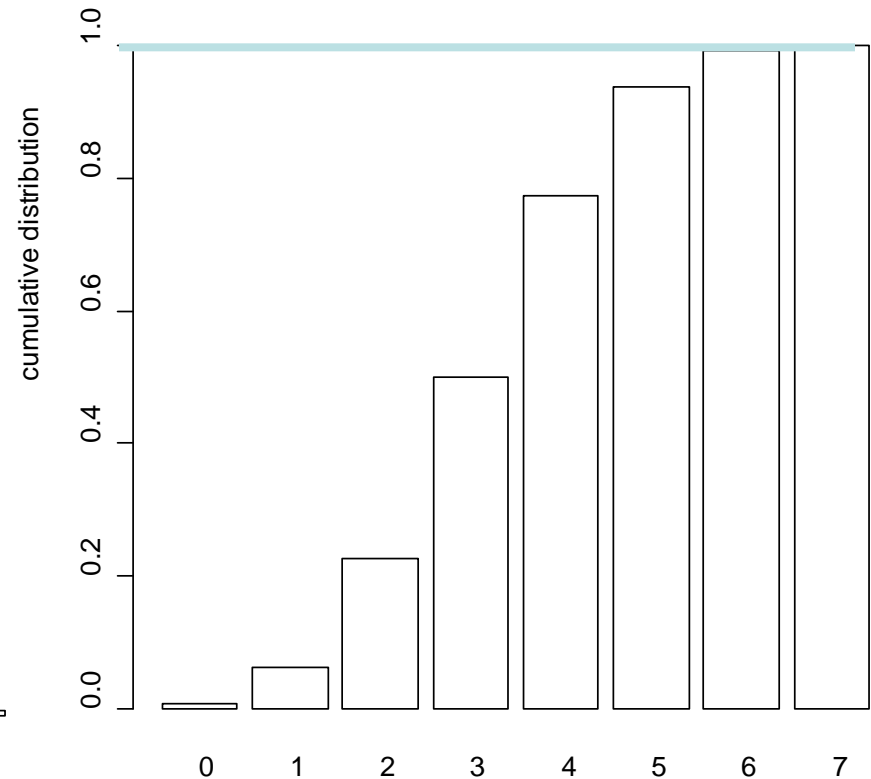




# Cumulative distribution



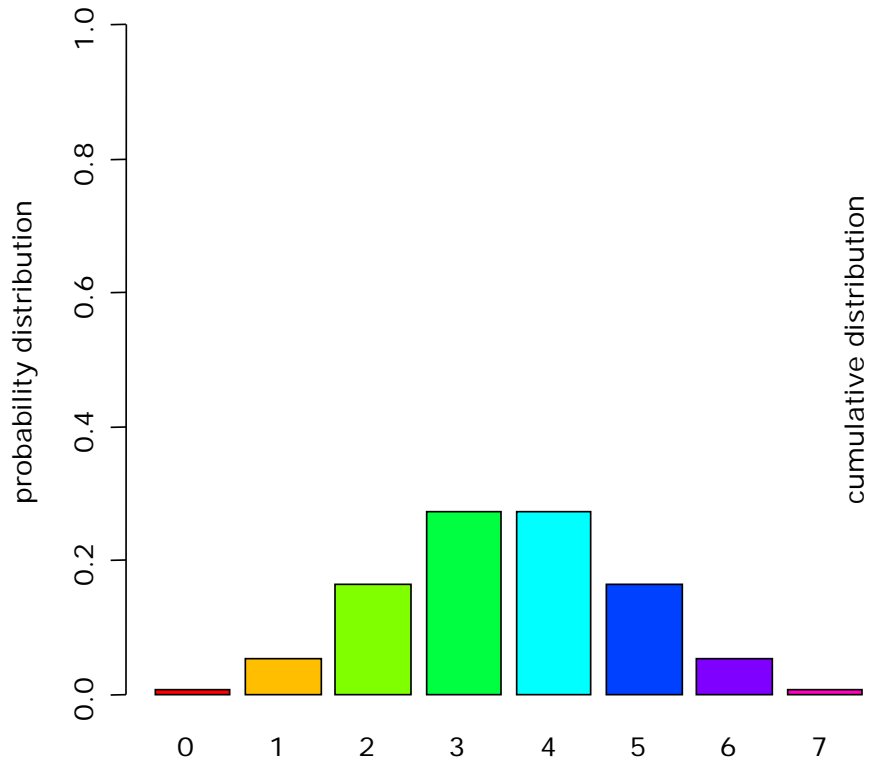
$P(X=x)$



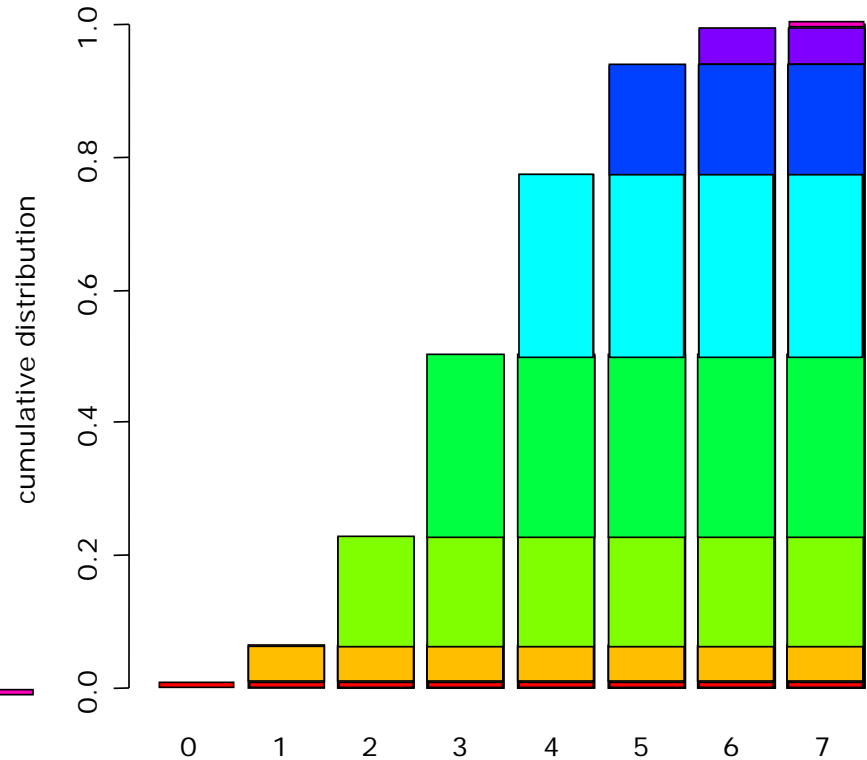
$P(X \leq x)$

> barplot(dbinom(0: 7, 7, 0.5), names.arg=0: 7)    > barplot(pbinom(0: 7, 7, 0.5), names.arg=0: 7)

# cumulative distribution



$P(X=x)$



$P(X \leq x)$

# example: surfers on a website

- Your site has a lot of visitors 45% of whom are female
- Out of the first 100 visitors, what's the probability that 55 or more are female.
- Hint:
  - probability that 55 or more are female is same as (1 - probability that 54 or less are female)
  - Or, probability that 55 or more are female is same as probability of 45 or less are male and 55% visitors are male
  - Use cumulative density function

# Solutions

- probability that 55 or more are female is same as (1 - probability that 54 or less are female)
- $P(X \geq 55) = 1 - P(X \leq 54)$   
 $= 1 - \text{pbinom}(54, 100, 0.45)$   
 $= 0.02839342$
- Or, probability that 55 or more are female is same as probability of 45 or less are male and 55% visitors are male
- $P(X \geq 55) = \text{pbinom}(100 - 55, 100, 1 - 0.45)$   
 $= \text{pbinom}(45, 100, 0.55)$   
 $= 0.02839342$

# Another way to calculate cumulative probabilities

- ?pbinom
- $P(X \leq x) = \text{pbinom}(x, \text{size}, \text{prob}, \text{lower.tail} = \text{T})$
- $P(X > x) = \text{pbinom}(x, \text{size}, \text{prob}, \text{lower.tail} = \text{F})$

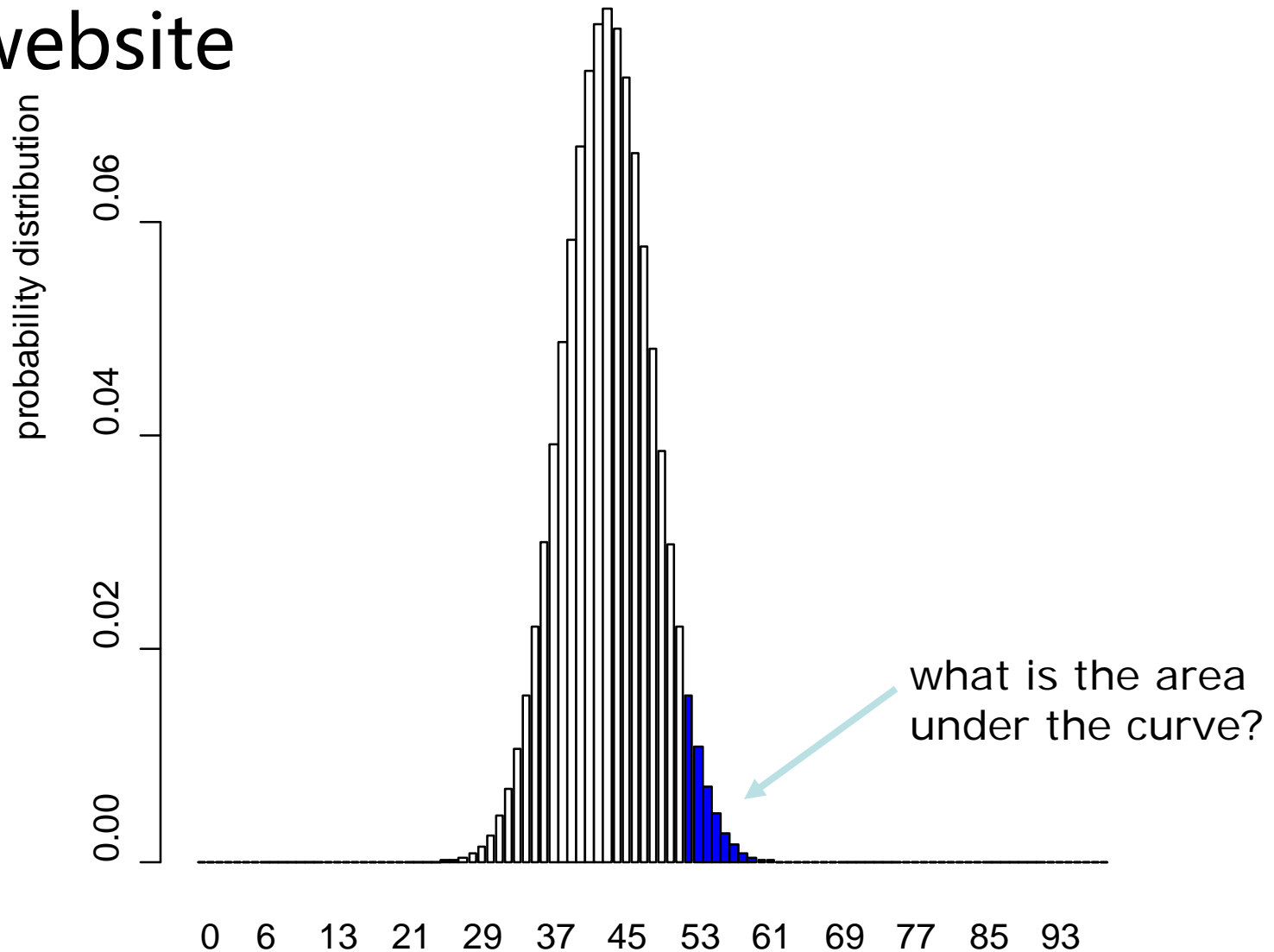
```
> 1-pbinom(54,100,0.45)
```

```
[1] 0.02839342
```

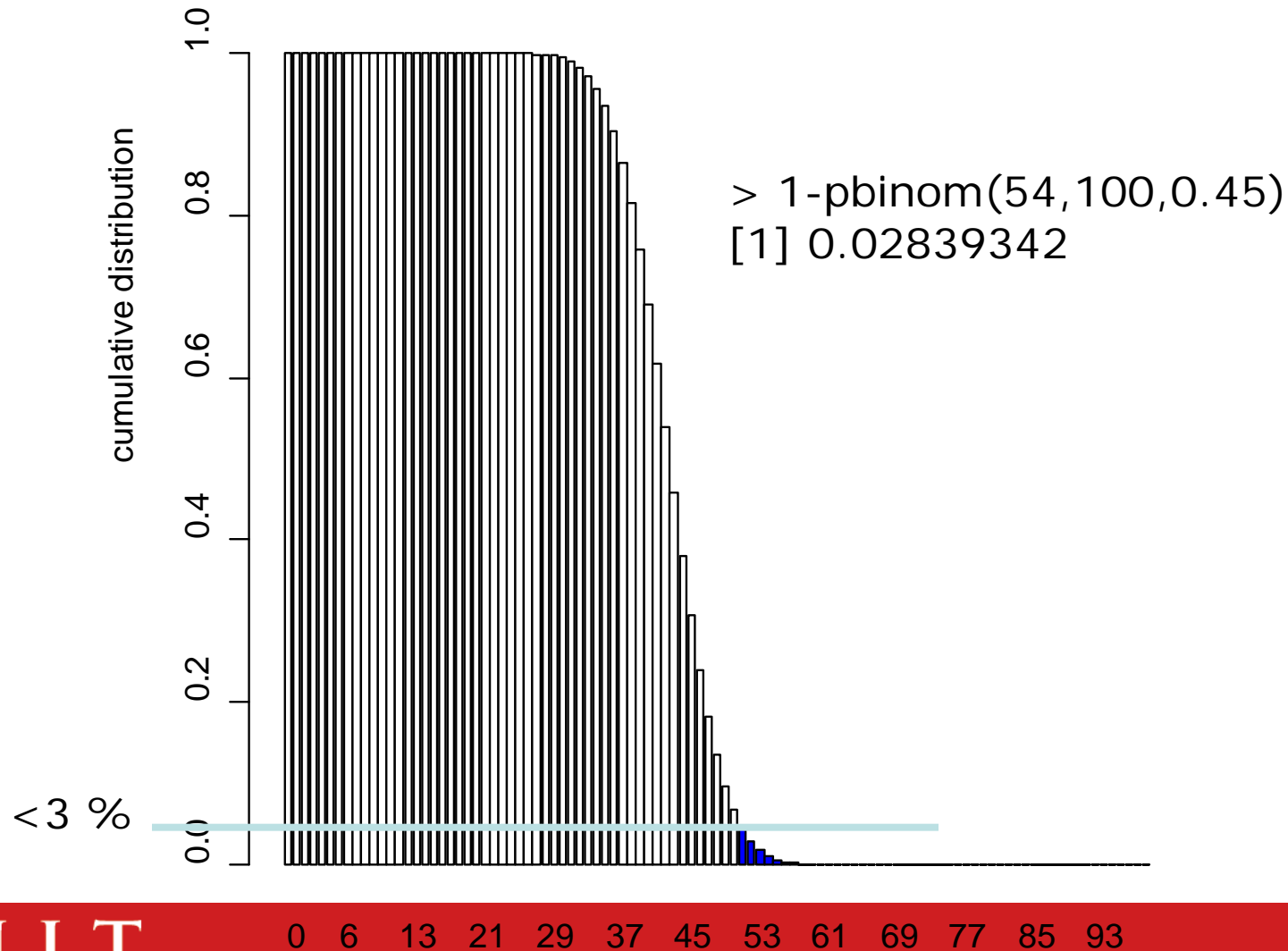
```
> pbinom(54,100,0.45,lower.tail=F)
```

```
[1] 0.02839342
```

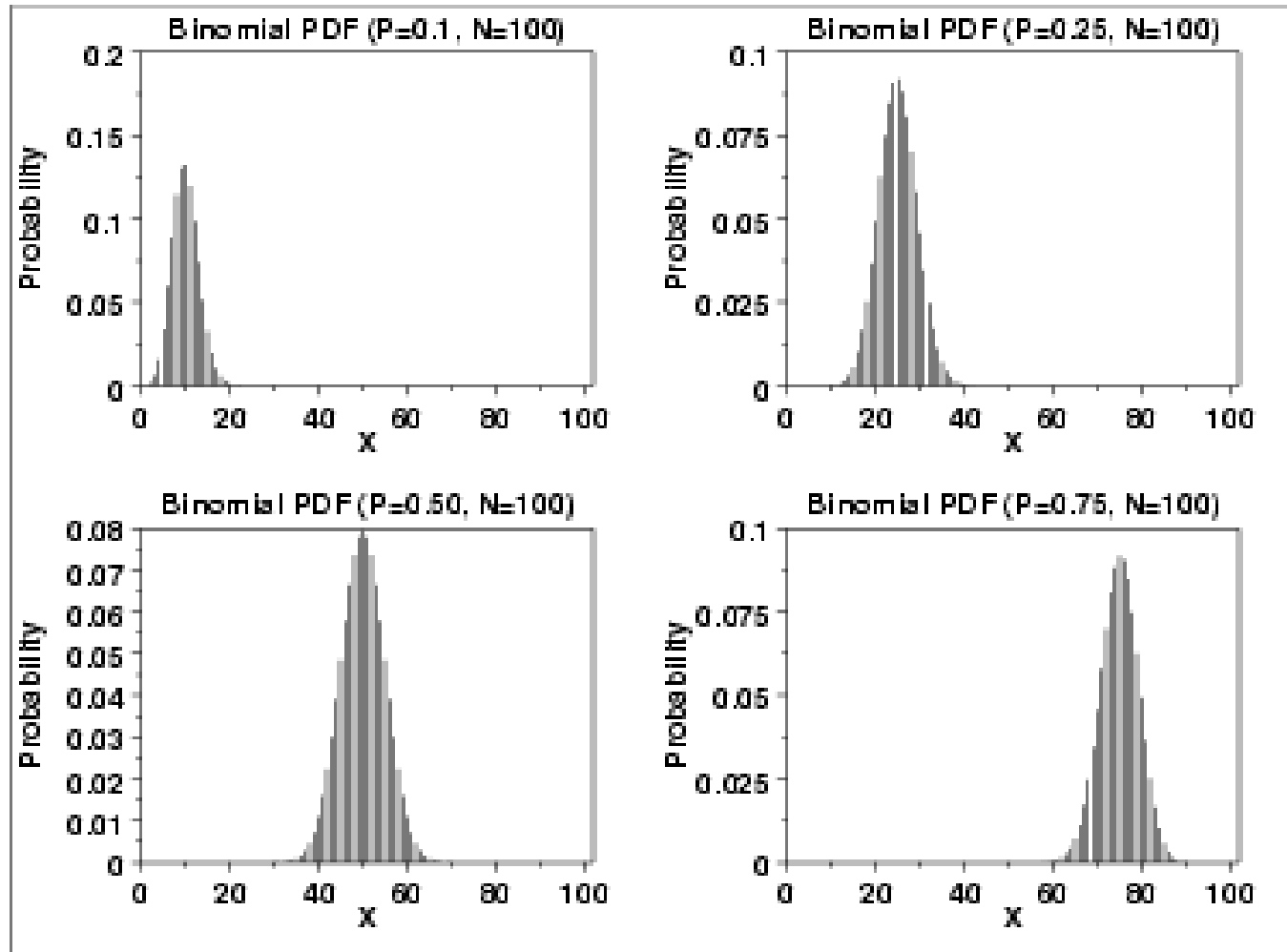
# Female surfers visiting a section of a website



# Cumulative distribution



# Plots of Binomial Distribution





# Another discrete distribution: hypergeometric

- Randomly draw  $n$  elements without replacement from a set of  $N$  elements,  $r$  of which are  $S'$  s (successes) and  $(N-r)$  of which are  $F'$  s (failures)
- hypergeometric random variable  $x$  is the number of  $S'$  s in the draw of  $n$  elements

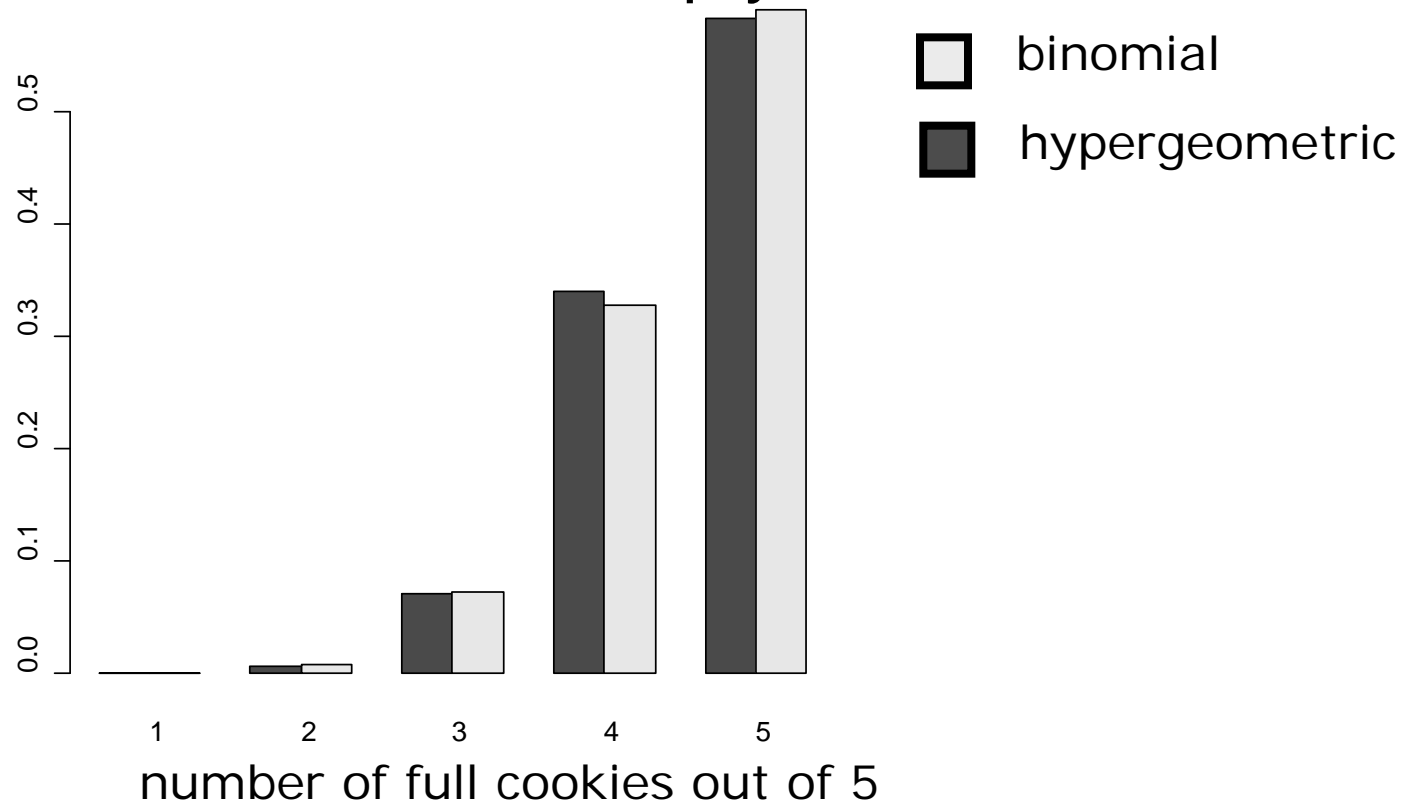
$$p(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

# hypergeometric example

- fortune cookies
- there are  $N = 20$  fortune cookies
- $r = 18$  have a fortune,  $N - r = 2$  are empty
- What is the probability that out of  $n = 5$  cookies,  $s = 5$  have a fortune (that is we don't notice that some cookies are empty)
- $> \text{dhyper}(5, 18, 2, 5)$
- [1] 0.5526316
- So there is a greater than 50% chance that we won't notice.

# hypergeometric and binomial

- When the population  $N$  is (very) big, whether one samples with or without replacement is pretty much the same
- 100 cookies, 10 of which are empty



# code aside

```
> x = 1:5  
> y1 = dhyper(1:5,90,10,5)  
> y2 = dbinom(1:5,5,0.9)  
> tmp = as.matrix(t(cbind(y1,y2)))  
> barplot(tmp,beside=T,names.arg=x)
```

hypergeometric probability

binomial probability

# Poisson distribution

- the number of times an event happened in a time interval
  - e.g. number of light bulbs burning out in a building in a year
  - number of people arriving in a queue per minute
  - number of users visited your website in an interval

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

□  $\lambda$  = mean # of events in a given interval

- $E[X] = \lambda, \text{Var}(X) = \lambda$

# Example: Poisson distribution

- You got a box of 1,000 widgets.
- The manufacturer says that the failure rate is 5 per box on average.
- Your box contains 10 defective widgets. What are the odds?

```
> ppois(9,5,lower.tail=F) #the CDF of  $X > 9$ 
```

```
[1] 0.03182806
```

```
> sum(dpois(10:1000, 5)) #sum PDF to get CDF
```

```
[1] 0.03182806
```

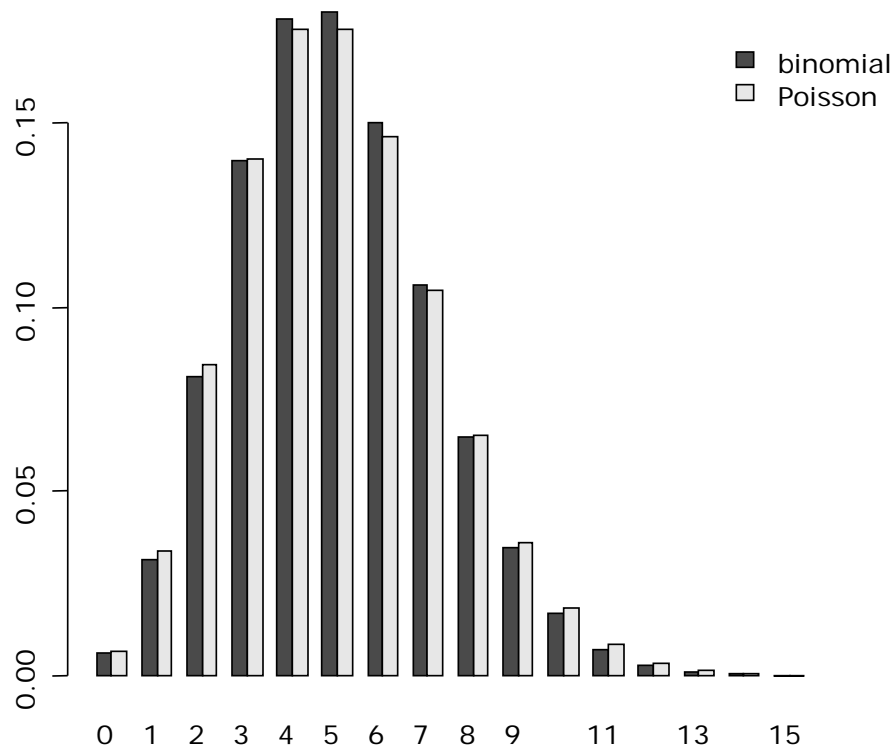
```
> dpois(10, 5) #PDF at 10
```

```
[1] 0.01813279
```

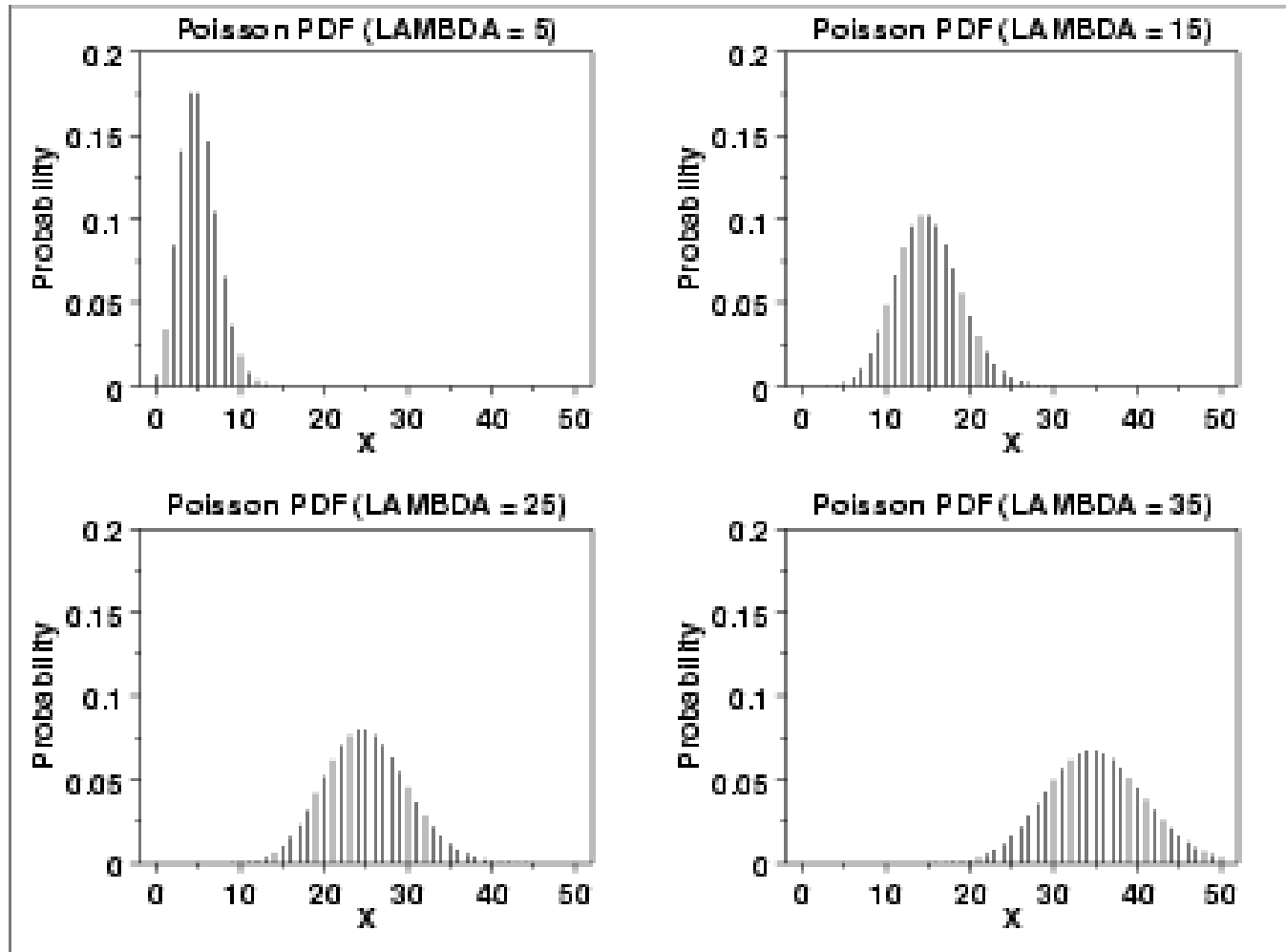
- The chance is 1.8%, maybe the manufacturer is not quite honest.
- Or the distribution is not Poisson?

# Poisson approximation to binomial

- If  $n$  is large (e.g.  $> 100$ ) and  $n \cdot p$  is moderate ( $p$  should be small) (e.g.  $< 10$ ), the Poisson is a good approximation to the binomial with  $\lambda = n \cdot p$



# Plots of Poisson Distribution





# Normal (Gaussian) Distribution

- Normal distribution (aka “bell curve” )
- fits many biological data well
  - e.g. height, weight
- serves as an **approximation** to binomial, hypergeometric, Poisson because of the Central Limit Theorem
- Well studied

# Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$

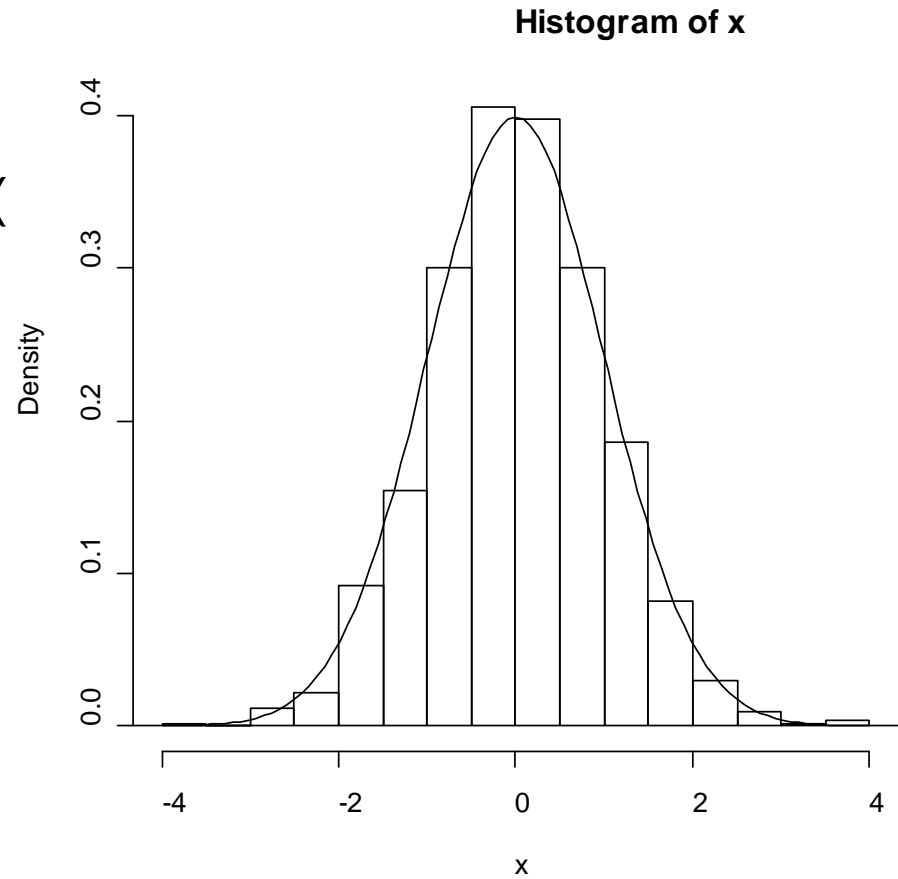
$$p_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$\Pr(a \leq X \leq b) = \int_a^b p_{\theta}(x) dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

- $E[X] = \mu, \text{Var}(X) = \sigma^2$
- If  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$ , and  $X_1, X_2$  are independent
  - $X = X_1 + X_2$  ?  $X \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ ,
  - $X = X_1 - X_2$  ?  $X \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ ,

# sampling from a normal distribution

```
x <- rnorm(1000)
h <- hist(x, plot=F)
ylim <-
  range(0,h$density,dnorm(
    0))
hist(x,freq=F,ylim=ylim)
curve(dnorm(x),add=T)
```



# Normal Approximation based on Central Limit Theorem

- Central Limit Theorem

- If  $x_i \sim \text{i.i.d}$  with  $(\mu, \sigma^2)$  and when  $n$  is large, then

$$(x_1 + \dots + x_n)/n \sim N(\mu, \sigma^2/n)$$

$$\text{Or } (x_1 + \dots + x_n) \sim N(n\mu, n\sigma^2)$$

- Example

- A population is evenly divided on an issue ( $p=0.5$ ). For a random sample of size 1000, what is the probability of having  $\geq 550$  in favor of it?

$n=1000$ ,  $x_i \sim \text{Bernoulli}(p=0.5)$ , i.e.  $E(x_i)=p$ ;  $V(x_i)=p(1-p)$

$$(x_1 + \dots + x_n) \sim \text{Binomial}(n=1000, p=0.5)$$

$$\Pr((x_1 + \dots + x_n) \geq 550) = 1 - \text{pbinom}(549, 1000, 0.5) = 0.000865$$

Normal Approximation:

$$(x_1 + \dots + x_n) \sim N(np, np(1-p)) = N(500, 250)$$

$$\Pr((x_1 + \dots + x_n) \geq 550) = 1 - \text{pnorm}(549, \text{mean}=500, \text{sd}=\sqrt{250}) = 0.000971$$

Why is the probability so small?

# d, p, q, and r functions in R

- In R, a set of functions have been implemented for each of almost all known distributions.
- `r<distname>(n,<parameters>)`
  - Possible distributions: binom, pois, hyper, norm, beta, chisq, f, gamma, t, unif, etc
- You find other characteristics of distributions as well
  - `d<dist>(x,<parameters>)`: density at x
  - `p<dist>(x,<parameters>)`: cumulative distribution function to x
  - `q<dist>(p,<parameters>)`: inverse cdf

# d, p, q, and r functions in R

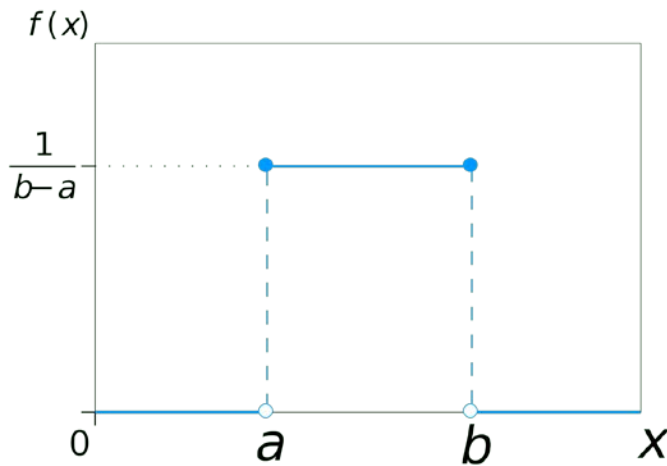
- p for "probability", the cumulative distribution function (c. d. f.)
- q for "quantile", the inverse c. d. f.
- d for "density", the density function (p. f. or p. d. f.)
- r for "random", a random variable having the specified distribution

Distribution	Functions			
<a href="#">Beta</a>	pbeta	qbeta	dbeta	rbeta
<a href="#">Binomial</a>	pbinom	qbinom	dbinom	rbinom
<a href="#">Cauchy</a>	pcauchy	qcauchy	dcauchy	rcauchy
<a href="#">Chi-Square</a>	pchisq	qchisq	dchisq	rchisq
<a href="#">Exponential</a>	pexp	qexp	dexp	rexp
<a href="#">E</a>	pf	qf	df	rf
<a href="#">Gamma</a>	pgamma	qgamma	dgamma	rgamma
<a href="#">Geometric</a>	pgeom	qgeom	dgeom	rgeom
<a href="#">Hypergeometric</a>	phyper	qhyper	dhyper	rhyper
<a href="#">Logistic</a>	plogis	qlogis	dlogis	rlogis
<a href="#">Log Normal</a>	plnorm	qlnorm	dlnorm	rlnorm
<a href="#">Negative Binomial</a>	pnbinom	qnbinom	dnbinom	rnbinom
<a href="#">Normal</a>	pnorm	qnorm	dnorm	rnorm
<a href="#">Poisson</a>	ppois	qpois	dpois	rpois
<a href="#">Student t</a>	pt	qt	dt	rt
<a href="#">Studentized Range</a>	ptukey	qtukey	dtukey	rtukey
<a href="#">Uniform</a>	punif	qunif	dunif	runif
<a href="#">Weibull</a>	pweibull	qweibull	dweibull	rweibull
<a href="#">Wilcoxon Rank Sum Statistic</a>	pwilcox	qwilcox	dwilcox	rwilcox
<a href="#">Wilcoxon Signed Rank Statistic</a>	psignrank	qsignrank	dsignrank	rsignrank

# Uniform Distribution

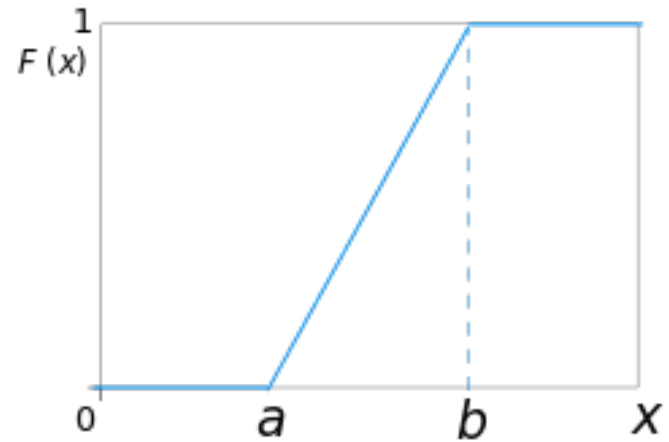
- The probability distribution function of the continuous uniform distribution is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$



Probability density function

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b] \\ 1 & \text{for } x > b \end{cases}$$



Cumulative distribution function

# Example: Uniform Distribution

- The uniform distr. On  $[a,b]$  has two parameter. The family name is *unif*. In R, the parameters are named *min* and *max*
- The uniform distribution has density
$$f(x) = 1/(max - min) \quad \text{for } min \leq x \leq max$$

```
> dunif(x=1, min=0, max=3)
```

```
[1] 0.3333333
```

```
> punif(q=2, min=0, max=3)
```

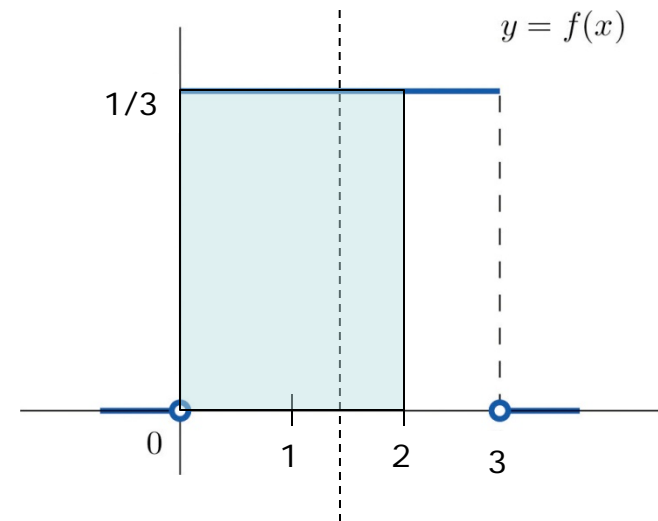
```
[1] 0.6666667
```

```
> qunif(p=0.5, min=0, max=3)
```

```
[1] 1.5
```

```
> runif(n=5, min=0, max=3)
```

```
[1] 2.7866852 1.9627136 0.9594195 2.6293273 1.6277597
```





# Bayes' Rule

- Given two events A and B and suppose that  $\Pr(A) > 0$ . Then

$$\Pr(B | A) = \frac{\Pr(AB)}{\Pr(A)} = \frac{\Pr(A | B) \Pr(B)}{\Pr(A)}$$

- Example:

$$\Pr(R) = 0.8$$

$\Pr(W R)$	R	$\neg R$
W	0.7	0.4
$\neg W$	0.3	0.6

R: It is a rainy day

W: The grass is wet

$$\Pr(R|W) = ?$$

# Bayes' Rule

$$\begin{aligned}P(R|W) &= P(RW)/P(W) \\&= P(WR)/P(W) \\&= P(WR)/[P(WR) + P(W\neg R)] \\&= P(W|R)P(R)/[P(W|R)P(R) + P(W|\neg R)P(\neg R)] \\&= 0.7*0.8/(0.7*0.8+0.4*0.2)\end{aligned}$$

# Summation (Integration) out tip

- Suppose that  $B_1, B_2, \dots, B_k$  form a partition of  $\Omega$ :  
 $B_1 \cup B_2 \cup \dots \cup B_k = \Omega$  and  $B_1, B_2, \dots, B_k$  are mutually exclusive

Suppose that  $\Pr(B_i) > 0$  and  $\Pr(A) > 0$ . Then

$$\Pr(B_i | A) = \frac{\Pr(A | B_i) \Pr(B_i)}{\Pr(A)}$$

# Summation (Integration) out tip

- Suppose that  $B_1, B_2, \dots, B_k$  form a partition of  $\Omega$ :  
 $B_1 \cup B_2 \cup \dots \cup B_k = \Omega$  and  $B_1, B_2, \dots, B_k$  are mutually exclusive

Suppose that  $\Pr(B_i) > 0$  and  $\Pr(A) > 0$ . Then

$$\begin{aligned}\Pr(B_i | A) &= \frac{\Pr(A | B_i) \Pr(B_i)}{\Pr(A)} \\ &= \frac{\Pr(A | B_i) \Pr(B_i)}{\sum_{j=1}^k \Pr(AB_j)}\end{aligned}$$

# Summation (Integration) out tip

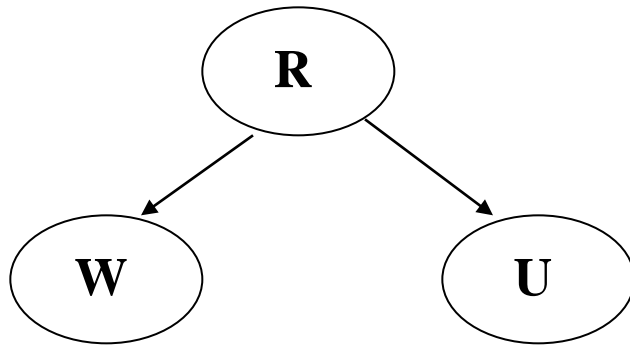
- Suppose that  $B_1, B_2, \dots, B_k$  form a partition of  $\Omega$ :  
 $B_1 \cup B_2 \cup \dots \cup B_k = \Omega$  and  $B_1, B_2, \dots, B_k$  are mutually exclusive

Suppose that  $\Pr(B_i) > 0$  and  $\Pr(A) > 0$ . Then

$$\begin{aligned}\Pr(B_i | A) &= \frac{\Pr(A | B_i) \Pr(B_i)}{\Pr(A)} \\ &= \frac{\Pr(A | B_i) \Pr(B_i)}{\sum_{j=1}^k \Pr(AB_j)} \\ &= \frac{\Pr(A | B_i) \Pr(B_i)}{\sum_{j=1}^k \Pr(B_j) \Pr(A | B_j)}\end{aligned}$$

**Key: Joint distribution!**

# Application of Bayes' Rule

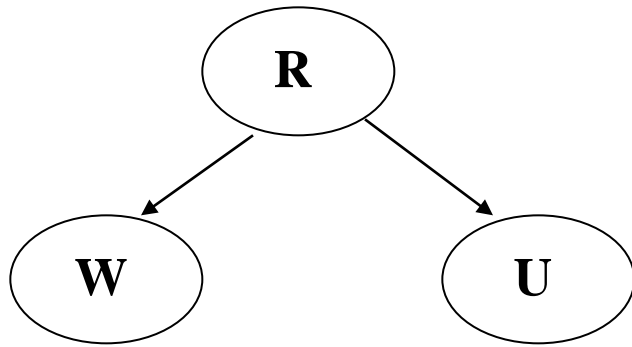


**R** It rains

**W** The grass is wet

**U** People bring umbrella

# A More Complicated Example



**R** It rains

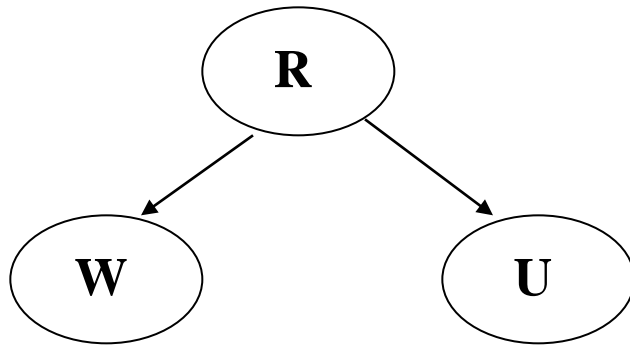
**W** The grass is wet

**U** People bring umbrella

$$\Pr(UW|R) = \Pr(U|R)\Pr(W|R)$$

$$\Pr(UW|\neg R) = \Pr(U|\neg R)\Pr(W|\neg R)$$

# A More Complicated Example



$$\Pr(R) = 0.8$$

**R** It rains

**W** The grass is wet

**U** People bring umbrella

$$\Pr(UW|R) = \Pr(U|R)\Pr(W|R)$$

$$\Pr(UW|\neg R) = \Pr(U|\neg R)\Pr(W|\neg R)$$

$\Pr(W R)$	R	$\neg R$
W	0.7	0.4
$\neg W$	0.3	0.6

$\Pr(U R)$	R	$\neg R$
U	0.9	0.2
$\neg U$	0.1	0.8

$$\Pr(U|W) = ?$$



# Solution:

$$\begin{aligned}P(UW) &= P(UW|R)P(R) + P(UW|\neg R)P(\neg R) \\&= P(U|R)P(W|R)P(R) + \\&\quad P(U|\neg R)P(W|\neg R)P(\neg R) \\&= 0.9 * 0.7 * 0.8 + 0.2 * 0.4 * 0.2 \\&= 0.52\end{aligned}$$

$$\begin{aligned}P(W) &= P(W|R)P(R) + P(W|\neg R)P(\neg R) \\&= 0.7 * 0.8 + 0.4 * 0.2 \\&= 0.64\end{aligned}$$

$$P(U|W) = P(UW)/P(W) = 0.52/0.64 = 0.8125$$

# Acknowledgments

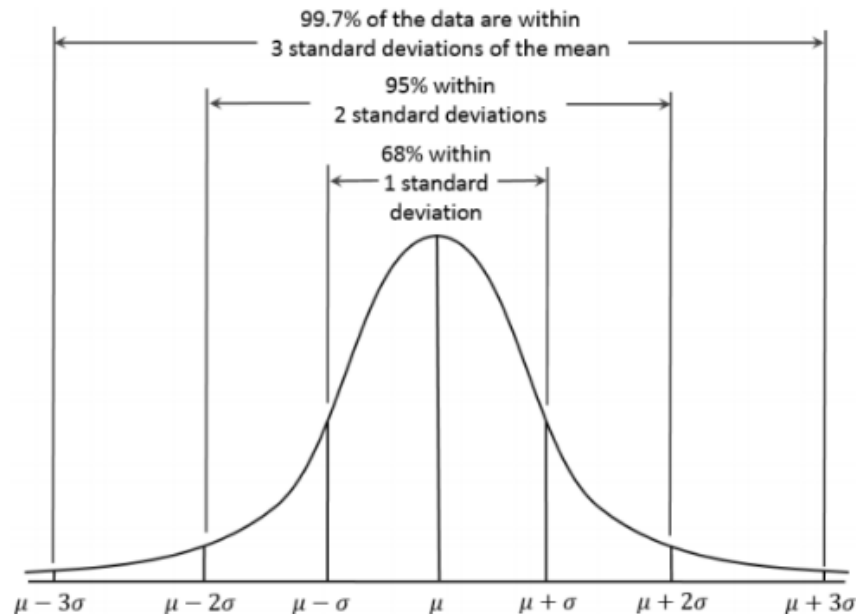
- Peter N. Belhumeur: for some of the slides adapted or modified from his lecture slides at Columbia University
- Rong Jin: for some of the slides adapted or modified from his lecture slides at Michigan State University
- Jeff Solka: for some of the slides adapted or modified from his lecture slides at George Mason University
- Brian Healy: for some of the slides adapted or modified from his lecture slides at Harvard University

# Homework

## Problem 1, Six Sigma

[https://en.wikipedia.org/wiki/Six\\_Sigma](https://en.wikipedia.org/wiki/Six_Sigma)

Six Sigma ( $6\sigma$ ) is a set of techniques and tools for process improvement. It was introduced by American engineer Bill Smith while working at Motorola in 1980. Jack Welch made it central to his business strategy at General Electric in 1995. A six sigma process is one in which **99.99966% of all opportunities to produce some feature of a part are statistically expected to be free of defects.**



1. Please write R code to verify the probabilities within  $\sigma$ ,  $2\sigma$  and  $3\sigma$  as shown in the above graph
2. Write R code to find out, within how many  $\sigma$ , the probability is 99.99966%. Is it really  $6\sigma$ ?

# Homework

## Problem 2, Job Search

You are spamming resumes to apply jobs. Each resume you send out has 1% chance of getting a job offer. You have sent 100 resumes. What's your chance to get at least one job offer?

1. Please solve the problem by math. Use R mathematical expression to get the result.
2. Analytic solution. Use R's probability functions to solve the problem.
3. Answer by simulation. Use sampling function to simulate the process and estimate the answer.
4. How many resumes in total do you have to spam so that you will have 90% chance to get at least one job offer?

## Problem 3, President Election Polls

Half of the population supports the president (i.e.,  $p=0.5$ ). For a random sample of size 1000, what is the probability of having  $\geq 600$  in support of the president?

1. Use binomial distribution
2. Use normal distribution as approximation.