# Lecture 2b: Probability Distribution Functions

Robert Horton

December 28, 2014

# Objectives

- describe what a probability distribution function represents, and what it can be used for.
- for the following probability distributions, say whether they are continuous or discrete, and describe the kinds of systems they can be used to model:
  - binomial
  - Poisson
  - normal
- call an R function to generate a vector of random numbers from a normal distribution with a given mean and standard deviation.
- describe the conditions under which the Normal distribution closely approximates the binomial or Poisson distributions.

# The Central Limit Theorem

*I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it.*

–Sir Francis Galton, 1889

# The Galton Box

FIG. 7.

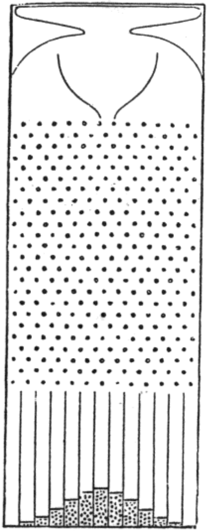


FIG. 8.

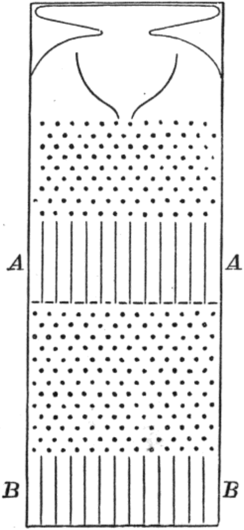
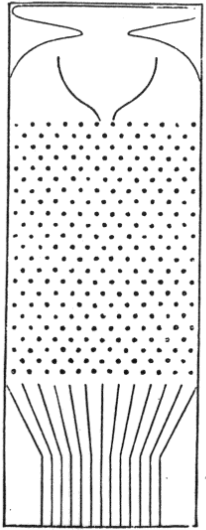
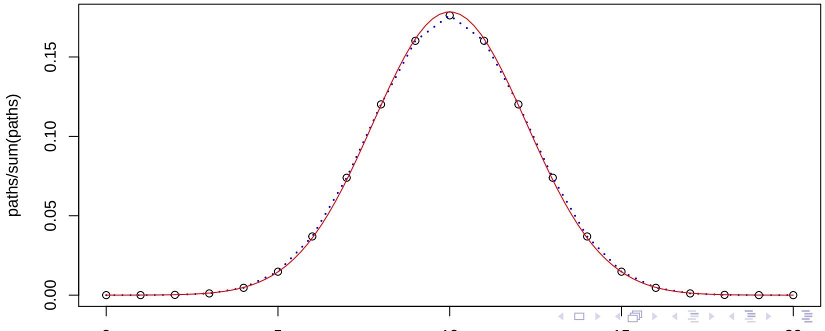



FIG. 9.



# Normal Approximation to the Binomial Distribution

```
n <- 20
k <- 0:n
paths <- choose(n,k)
plot(k, paths/sum(paths))
p <- 0.5
lines(k, dbinom(k, n, p), col="blue", lty=3, lwd=2)
x <- seq(0, n, length=100)
lines(x, dnorm(x, mean=n * p, sd=sqrt(n * p * (1-p))), col=
```



# DeMoivre - Laplace Theorem

de Moivre, A. *The Doctrine of Chances*, 1738

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

For the biomial distribution:

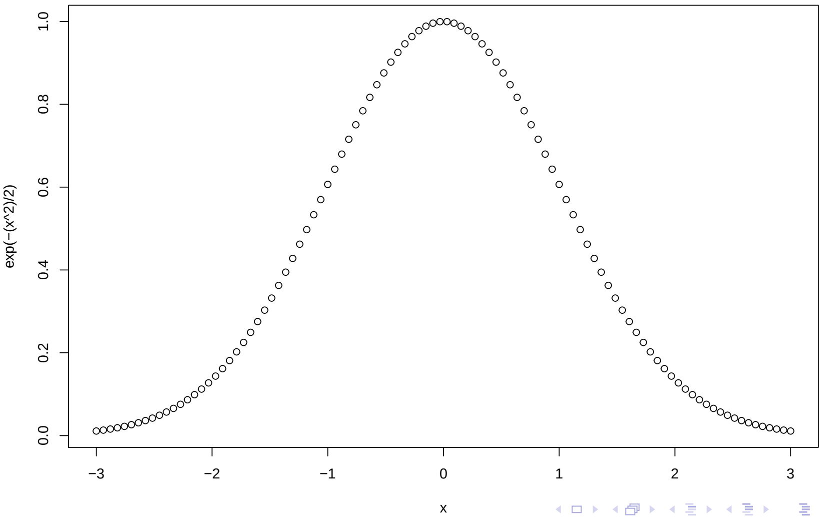
$$\mu = np$$

$$\sigma = \sqrt{npq}$$

where q = (1-p)

# Bell Curves

```
x <- seq(-3, 3, length=100)
plot(x, exp(-(x^2)/2))
```



# The Bell Curve as a Probability Distribution Function

```
integrate(function(x) exp(-(x^2)/2), -Inf, Inf)
```

```
## 2.506628 with absolute error < 0.00023
```

```
sqrt(2 * pi)
```

```
## [1] 2.506628
```

# The Normal Distribution

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

```
normal_pdf <- function(x, mu=0, sigma=1)
    1/(sigma * sqrt(2*pi)) * exp(-(x - mu)^2/(2*sigma^2))

integrate(normal_pdf, -Inf, Inf)
```

```
## 1 with absolute error < 9.4e-05
```

# The Normal Distribution

```
library(manipulate)
x <- seq(-20, 20, by=0.1)
manipulate(plot(x, dnorm(x, mean=m, sd=s), type="l", col="
           m = slider(-20.0, 20.0, initial=0),
           s = slider(0.01, 10, initial=1))
```

# The Normal Distribution

```
x <- rnorm(1e5)
hist(x)
```

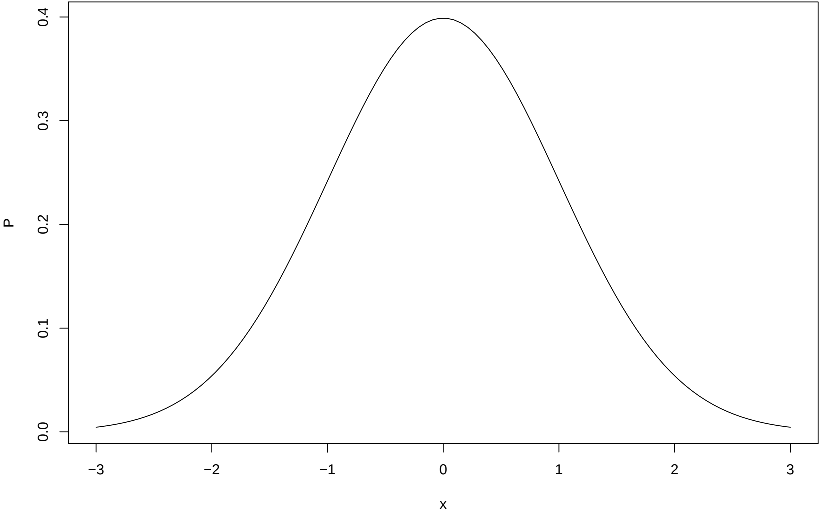
```
dfrm <- data.frame(
  x=c(x, x * 3 + 10),
  distro=rep(c("Normal", "m10sd3"), each=length(x))
)
require("ggplot2")
```

```
## Loading required package: ggplot2
```

```
g <- ggplot(dfrm, aes(x=x, col=distro)) + geom_density()
g
```

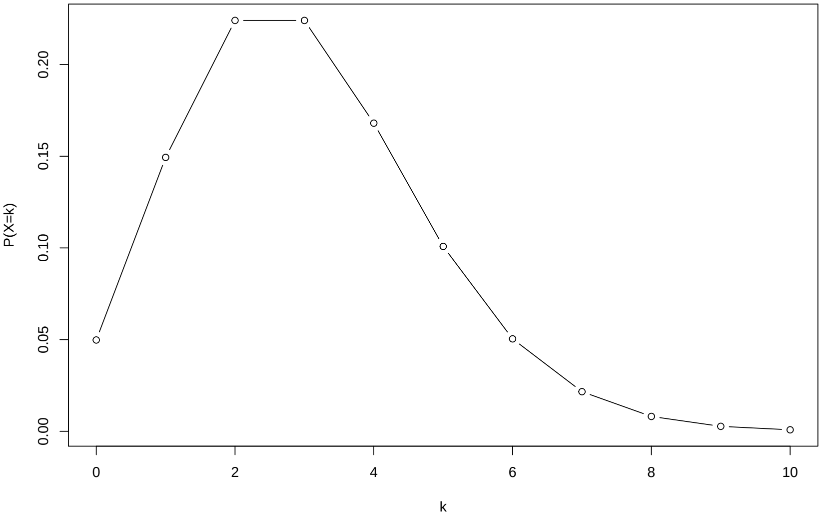
# Review Questions

Name this distribution:



# Review Questions

Which distribution might this be?:



# Review Questions

- Name two commonly used probability density functions (continuous)
- Name two commonly used probability mass functions (discrete)

# Assignment