Probability and Random variables

Partially based on IPSUR notes

Sample Spaces

For a random experiment E, the set of all possible outcomes of E is called the sample space and is denoted by the letter S . For a coin-toss experiment, S would be the results *Head* and *Tail*.
It may represented by $S = \{H, T\}$. Formally, the performance of a random experiment is the unpredictable selection of an outcome in S.

```
library(prob)
tosscoin(3)

# 6-sided die
rolldie(2)

# Cards
cards()
```

Events and Probability Functions I

- An event A is a subset of the sample space After the performance of a random experiment E. We say that the event A occurred if the experiment's outcome belongs to A.
- We say that a bunch of events $A_1, A_2, A_3 \ldots$ are mutually exclusive or disjoint if $A_i \cap A_j = \varnothing$ for any distinct pair $A_i \neq A_j$.
- A probability function is a rule that associates with each event A of the sample space a unique number $P(A) = p$, called the probability of A. Any probability function P satisfies the three *Kolmogorov Axioms*:

Events and Probability Functions II

- The probability of an event should never be negative. Since the sample space contains all possible outcomes, its probability should be one. Finally, for a sequence of disjoint events (sets that do not overlap), their total probability (measure) should equal the sum of its parts.

```
# Equally likely model
tosscoin(3, makespace=TRUE)
probspace(rolldie(2))

# Not equal probabilities
iidspace(c("H","T"), ntrials=3, probs=c(0.7, 0.3))
```

Conditional Probability

- **Definition:** The conditional probability of B given A, denoted $P(B|A)$, is defined by

$$P(B|A) = \frac{P(B \cap A)}{p(A)}$$

if $P(A) > 0$. When $P(A) = 0$, the theory forms the foundation for the study of stochastic processes.

```
S = rolldie(2, makespace=TRUE)
S
A = subset(S, X1 == X2)
B = subset(S, X1 + X2 >= 8)
prob(A, given=B)
prob(B, given=A)
```

Independence

- **Definition**: Events A and B are said to be independent if

$$P(A \cap B) = P(A)P(B)$$

  Otherwise, the events are said to be dependent.

- **Example**: Toss ten coins. What is the probability of observing at least one Head?

```
S = tosscoin(10, makespace=TRUE)
A = subset(S, isrep(S, vals="T", nrep=10))
1 - prob(A)
```

Bayes' Rule I

The Bayes' Rule allows us to update our probabilities when new information becomes available:
Let $B_1, B_2, \dots B_n$ be mutually exclusive and exhaustive and let A be an event with $P(A) > 0$. Then

$$P\left(B_k|A\right) = \frac{P\left(A|B_k\right) \cdot P\left(B_k\right)}{\sum_{i=1}^{n} P\left(A|B_i\right).P\left(B_i\right)}$$

for $k = 1, 2, \dots, n$.

Bayes' Rule II

Example

In this problem, there are three assistants working at a company: Moe, Larry, and Curly.
Their primary job duty is to file paperwork in the filing cabinet when papers become available. The three assistants have different work schedules:

	Moe	Larry	Curly
Workload	60%	30%	10%

That is, Moe works 60% of the time, Larry works 30% of the time, and Curly does the remaining 10%, and they file documents at approximately the same speed. Suppose a person were to select one of the documents from the cabinet at random.

Bayes' Rule II

Example.
Let M be the event M $= \{$Moe filed the document$\}$ and let L and C be the events that Larry and Curly, respectively, filed the document.
In the absence of additional information, reasonable prior probabilities would just be

$$P(M) = 0.60 \quad P(L) = 0.30 \quad P(C) = 0.10$$

Now, the boss comes in one day, opens up the file cabinet, and selects a file at random. The boss discovers that the file has been misplaced. The question is: who misplaced the file?

Bayes' Rule II

Example
The boss has information about Moe, Larry, and Curly's filing accuracy in the past (due to historical performance evaluations). The performance information may be represented by the following table:

	Moe	Larry	Curly
Misfile Rate	0.003	0.007	0.010

In other words, on the average, Moe misfiles 0.3% of the documents he is supposed to file and so on.

Bayes' Rule II

Example

We store the prior probabilities and the likelihoods in vectors and we apply the Bayes' Rule directly.

```
prior = c(0.6, 0.3, 0.1)
like = c(0.003, 0.007, 0.01)
post = prior * like
post/sum(post)
[1] 0.3673469 0.4285714 0.2040816
```

The conclusion: Larry probably misplaced the file...

Random Variables

We conduct a random experiment E and after learning the outcome ω in S we calculate a number X. That is, to each outcome ω in the sample space we associate a number $X(\omega) = x$.

Definition: A random variable X is a function $X : S \to \mathbb{R}$ that associates to each outcome $\omega \in S$ exactly one number $X(\omega) = x$.

Example: Let E be the experiment of flipping a coin twice. Now define the random variable $X =$ the number of heads. That is, for example, $X(HH) = 2$, while $X(HT) = 1$. We may make a table of the possibilities

$\omega \in S$	HH	HT	TH	TT
$X(\omega) = x$	2	1	1	0

Random Variables

Example: let us roll a die three times, and let us define the random variables

$$
\begin{aligned}
U &= X_1 - X_2 + X_3 \\
V &= \max(X_1, X_2, X_3) \\
W &= X_1 + X_2 + X_3
\end{aligned}
$$

```
S = rolldie(3,  makespace=TRUE)
S = addrv(S, U = X1 - X2 + X3)
S = addrv(S, FUN=max, invars=c("X1","X2","X3"),
name="V")
S = addrv(S, FUN=sum, invars=c("X1","X2","X3"),
name="W")
S

prob(S, U > 6)
prob(S, U + W - V > 10)
```

Discrete Distributions

Discrete random variables are characterized by their supports which take the form

$$S_X = \{u_1, u_2, \ldots\}$$

Every discrete random variable X has associated with it a probability mass function (PMF) $f_X : S_X \to [0;1]$ defined by

$$f_X(x) = P(X = x)$$

for $x \in S_X$.
Mean and Variance:

$$\begin{aligned}
\mu &= E(X) = \sum_{x \in S} x f_X(x) \\
\sigma^2 &= \sum_{x \in S} (x - \mu)^2 f_X(x)
\end{aligned}$$

Discrete Distributions

Example:

```
x = c(0,1,2,3)
f = c(1/8, 3/8, 3/8, 1/8)

mu = sum(x * f); mu
sigma2 = sum((x-mu)^2 * f); sigma2
sigma = sqrt(sigma2); sigma

# Using an specific library
library(distrEx)
X = DiscreteDistribution(supp=0:3, prob=c(1,3,3,1)/8)
E(X); var(X); sd(X)
```

Binomial Distribution

The binomial distribution is based on a Bernoulli trial, which is a random experiment in which there are only two possible outcomes: success (S) and failure (F). We conduct the Bernoulli trial and let

$$X = \begin{cases} 1 \text{ if the outcomes is } S \\ 0 \text{ if the outcomes is } F \end{cases}$$

The probability function is

$$f_X(x) = p^x(1-p)^{1-x}$$

for $x = 0, 1$.
The Binomial model has three defining properties:

- Bernoulli trials are conducted n times,
- the trials are independent,
- the probability of success p does not change between trials.

Binomial Distribution

The probability function is

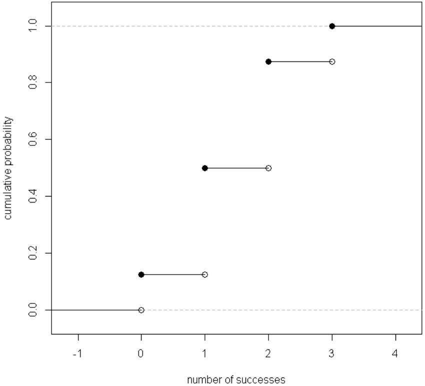
$$f_X(x) = \binom{n}{x} p^x (1-p)^{1-x}$$

for $x = 0, 1, 2, \ldots, n$

```
A <- data.frame(Pr=dbinom(0:3, size=3, prob=0.5))
rownames(A) <- 0:3
A

plot(0, xlim=c(-1.2, 4.2), ylim=c(-0.04, 1.04),
type="n", xlab="number of successes",
ylab="cumulative probability")
abline(h=c(0,1), lty=2, col="grey")
```

Binomial Distribution

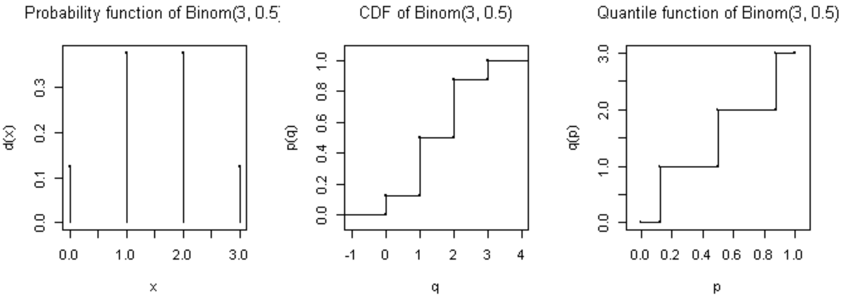
```
lines(stepfun(0:3, pbinom(-1:3, size=3, prob=0.5)),
verticals=FALSE, do.p=FALSE)
points(0:3, pbinom(0:3, size=3, prob=0.5), pch=16,
cex=1.2)
points(0:3, pbinom(-1:2, size=3, prob=0.5), pch=1,
cex=1.2)
```



Binomial Distribution

We can use the library `distr`

```
library(distr)
X = Binom(size=3, prob=1/2)

d(X)(1) # pmf of X evaluated at x=1
p(X)(2) # cdf of X evaluated at x=2
op <- par(pty="s") # square plotting region
plot(X, cex=0.2)
par(op)
```



Binomial Distribution

In general,

Given $X \sim$ `dbinom(size=n, prob=p)`

How to do	with `stats` (*default*)	with `distr`
PMF: $P(X = x)$	`dbinom(x, size=n, prob=p)`	`d(X)(x)`
CDF: $P(X \leq x)$	`pbinom(x, size=n, prob=p)`	`p(X)(x)`
Simulate k variates	`rbinom(k, size=n, prob=p)`	`r(X)(k)`

For using the library `distr` we need to write previously

`X = Binom(size=n, prob=p)`

```
# Example
X <- Binom(size=3, prob=0.45)
library(distrEx)
E(X)
E(3*X + 4)
```

The Poisson Distribution

This is a distribution associated with "rare events", like traffic accidents, typing errors, or customers arriving in a bank.
Let λ be the average number of events, then,

$$f_X(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

The associated R functions are: `dpois(x, lambda)`, `ppois(x, lambda)`, `qpois(x, lambda)`, `rpois(n, lambda)` which give the PMF, CDF, quantile function, and simulate random variates, respectively.

Example: Suppose $Y \sim \text{Pois}(\texttt{lambda} = 50)$, compute
$P(48 \leq Y \leq 50) = P(X \leq 50) - P(X \leq 47)$.

```
diff(ppois(c(47, 50), lambda=50))
```

The Empirical Distribution

Do an experiment n times, and observe n values $x_1, x_2, \ldots, x_n$ of a random variable X. The empirical cumulative distribution function F_n (written *ECDF*) is the probability distribution that places probability mass $1/n$ on each of the values $x_1, x_2, \ldots, x_n$. The empirical PMF takes the form

$$f_X(x) = \frac{1}{n}$$

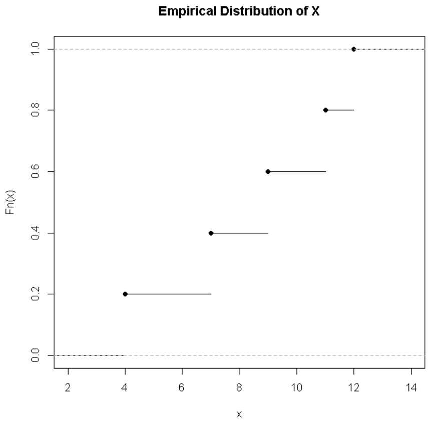
for $x \in \{x_1, x_2, \ldots x_n\}$.
Mean and variance are

$$\begin{aligned}
\mu &= \sum_{i=1}^{n} x_i \cdot \frac{1}{n} = \overline{x} \\
\sigma^2 &= \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \frac{1}{n}
\end{aligned}$$

The Empirical Distribution

With R: The graph is of a right-continuous function with jumps exactly at the locations stored in x:

```
x = c(4, 7, 9, 11, 12)
ecdf(x)
plot(ecdf(x), main='Empirical Distribution of X')
```



The Empirical Distribution

A function can be defined to compute the empirical PDF in each point:

```
epdf = function(x,t){
sum(x %in% t)/length(x)
}
x = c(0,0,1)
epdf(x,0) # should be 2/3
```

To simulate from the empirical distribution supported on the vector x, we use the `sample` function.

```
x = c(0, 0, 1)
sample(x, size=7, replace=TRUE)
```

Continuous Random Variables

Continuous random variables have supports like

$$S_X = [a, b] \text{ or } (a, b),$$

or unions of intervals of the above form. For example,

- the height or weight of an individual,
- physical measurements such as the length or size of an object, and
- durations of time (usually).

Every continuous random variable X has a *probability density function* (PDF) denoted f_X associated with it.

Continuous Random Variables

It satisfies three basic properties:

1. $f_X(x) > 0$ for $x \in S_X$,
2. $\int_{x \in S_X} f_X(x)dx = 1$, and
3. $P(X \in A) = \int_{x \in A} f_X(x)dx$, for an event $A \subset S_X$.

The mean μ, also known as $E(X)$:

$$\mu = E(X) = \int_{x \in S} x f_X(x)dx,$$

provided $\int_S |x| f(x)dx$ is finite. The variance is

$$\sigma^2 = E(X-\mu)^2 = \int_{x \in S} (x-\mu)^2 f_X(x)dx,$$

or alternatively $\sigma^2 = E(X^2) - (E(X))^2$.

Continuous Random Variables

Example: Let X have PDF $f(x) = 3x^2$, $0 < x < 1$ and find $P(0.14 \leq X \leq 0.71)$, $E(X)$ and $Var(X)$.

```
f = function(x) {3*x^2}
integrate(f, lower=0.14, upper=0.71)

# With library distr
library(distr)

X = AbscontDistribution(d=f, low1=0, up1=1)
p(X)(0.71) - p(X)(0.14)

# With library distrEx
# Mean and Variance
library(distrEx)

E(X)
var(X)
```

Normal Distribution

We say that X has a *normal distribution* if it has PDF

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

We write $X \sim N(\mu, \sigma)$, and the associated R function is `dnorm(x, mean=0, sd=1)`

The familiar bell-shaped curve, the normal distribution is also known as the *Gaussian distribution*
This distribution is by far the most important distribution, continuous or discrete. The normal model appears in the theory of all sorts of phenomena.

```
curve(dnorm(x), from=-5, to=5, ylab="y",
main="Normal Density")
# Some quantiles
qnorm(c(0.025, 0.01, 0.005), lower.tail=FALSE)
```

Functions of Continuous Random Variables

Let X have PDF f_X and let g be a function which is one-to-one with a differentiable inverse g^{-1}.
Then, the PDF of $U = g(X)$ is given by

$$f_U(u) = f_X\left[g^{-1}(u)\right] \left|\frac{d}{du}g^{-1}(u)\right|.$$

It is better to write in the intuitive form

$$f_U(u) = f_X(x)\left|\frac{dx}{du}\right|.$$

```
library(distr)
X <- Norm(mean=0, sd=1)
Y <- 4 - 3*X
p(Y)(0.5)
plot(Y)
W <- sin(exp(X) + 27)
p(W)(0.5)
plot(W)
```

Other Important Distributions: Uniform Distribution

A random variable X with the continuous uniform distribution on the interval (a, b) has PDF

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b.$$

The associated R function is $\text{dunif}(\texttt{min} = a, \texttt{max} = b)$.
It is used to model experiments whose outcome is an interval of numbers that are equally likely in the sense that any two intervals of equal length in the support have the same probability associated with them.
The mean of $X \sim \text{unif}(\texttt{min} = a, \texttt{max} = b)$ is

$$\mu = E(X) = \frac{b+a}{2}$$

Other Important Distributions: Exponential Distribution

We say that X has an *exponential distribution* and write $X \sim \exp(\texttt{rate} = \lambda)$. It is closely related to the Poisson distribution.
If customers arrive at a store, according to exponential distibuted times with rate λ, and if Y counts the number of customers that arrive in the time interval $[0, t)$, then $Y \sim \text{Pois}(\texttt{lambda} = \lambda t)$.

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$$

The associated R functions are: `dexp(x, rate)`, `pexp(x, rate)`, `qexp(x, rate)`, `rexp(n, rate)` which give the PMF, CDF, quantile function, and simulate random variates, respectively.

```
curve(dexp(x, rate=2), from=0, to=5, ylab="f(x)",
main="Exponential Distribution")
```

Other Important Distributions: Chi square Distribution

A random variable X with PDF

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, \quad x > 0,$$

is said to have a *chi-square distribution* with p *degrees of freedom*. The associated R functions are `dchisq(x, df)`, `pchisq`, `qchisq`, and `rchisq`, which give the PDF, CDF, quantile function, and simulate random variates, respectively.

```
curve(dchisq(x, df=3), from=0, to=20, ylab="f(x)")

ind <- c(4, 5, 10, 15)
for (i in ind) curve(dchisq(x, df=i), 0, 20, add=TRUE)
```

Other Important Distributions: t Student Distribution

A random variable X with PDF

$$f_X(x) = \frac{\Gamma\left[(r+1)/2\right]}{\sqrt{r\pi}\,\Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}, \quad -\infty < x < \infty$$

is said to have *Student's t* distribution with r *degrees of freedom*. The associated R functions are `dt(x, df)`, `pt(x, df)`, `qt(x, df)` and `rt(n, df)`.

```
curve(dt(x, df=50), from=-5, to=5, xlab="y",
ylab="f(x)", col="yellow")

curve(dt(x, df=3), from=-5, add=TRUE, col="blue")
```

Other Important Distributions: F Distribution

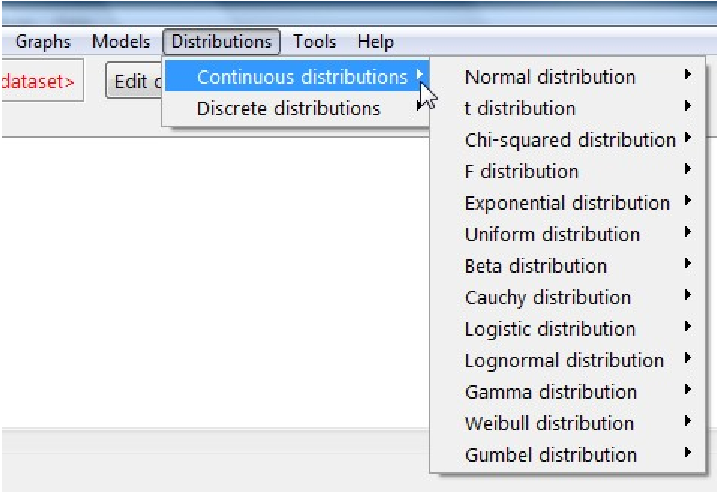
A random variable X with p.d.f.

$$f_X(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2}, \quad x > 0.$$

is said to have an F distribution with (m, n) degrees of freedom. The associated R functions are `df(x, df1, df2)`, `pf(x, df1, df2)`, `qf(x, df1, df2)` and `rf(n, df1, df2)`.

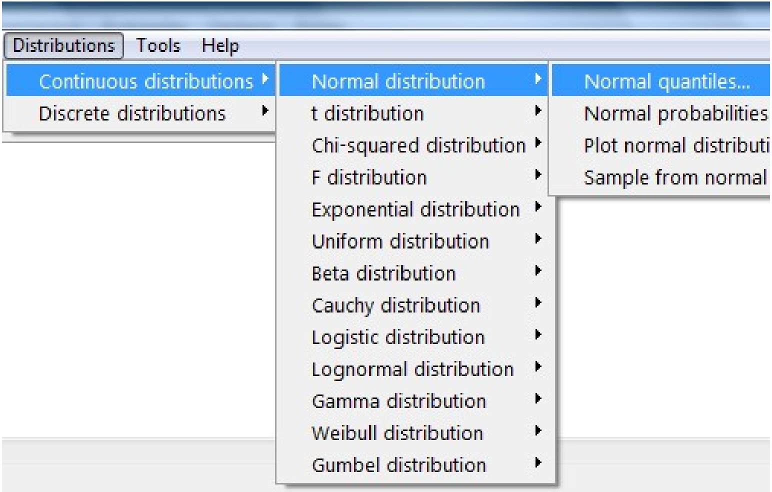
```
X11()
curve(df(x, df1=3, df2=20), from=0, to=20,
xlab="y", ylab="f(x)", col="yellow")

curve(df(x, df1=10, df2=2), from=0, to=20,
add=TRUE, col="blue")
```

Rcmdr: Distributions of continuous and discrete random variables



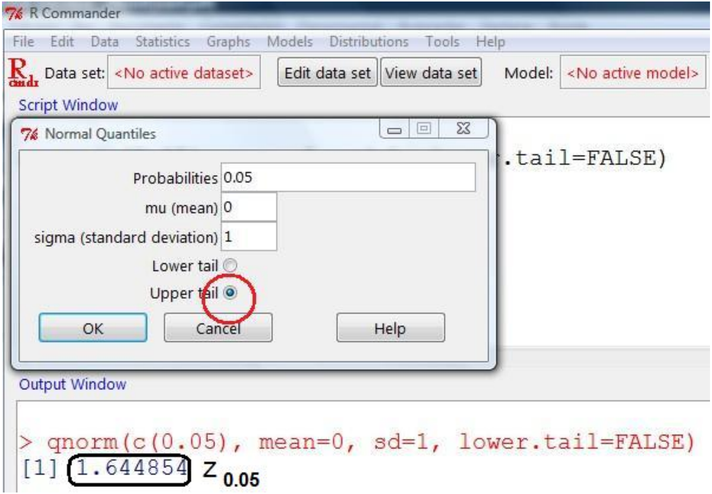
Finding (upper) quantiles of a distribution

- Find (upper) α quantile of the standard normal distribution, i.e., find a number z_α such that $P(Z > z_\alpha) = \alpha$ for $\alpha = 0.05$



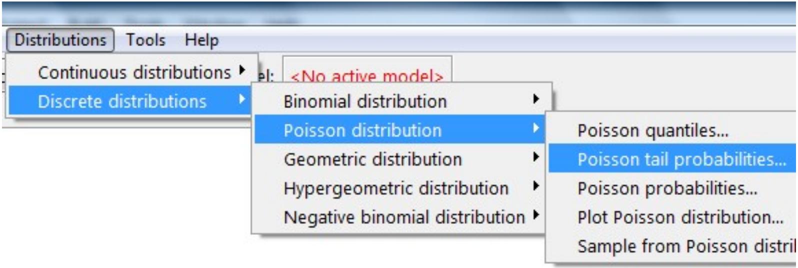
Finding (upper) quantiles of a distribution cont.

- $P(Z > z_{0.05}) = 0.05$ is satisfied by $z_{0.05} = 1.64$



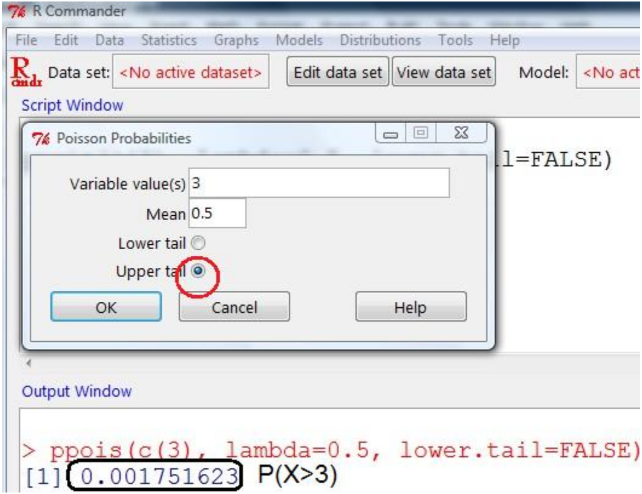
Finding probabilities of a distribution

- For $X \sim Poisson(\lambda = 2)$ $\boxed{E[X] = 1/\lambda = 0.5}$ find $P(X > 3)$
- Upper probabilities are with strict $>$ inequality
- Lower probabilities are with $\leq$



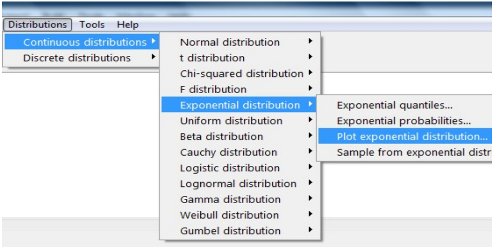
Finding probabilities of a distribution cont.

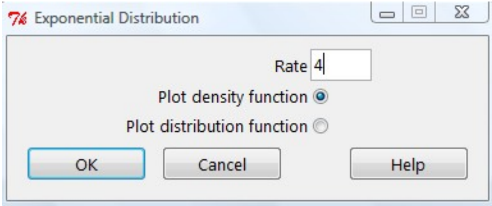
- For $X \sim Poisson(\lambda = 2)$, $P(X > 3) = 0.001752$
- To find $P(X \geq 3) = P(X > 2)$, set 2 in Variable value(s)



Plotting probability density function, pdf

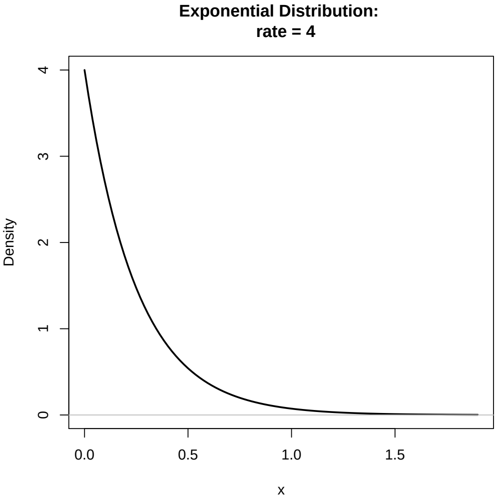
- For $X \sim Exp(\alpha = 4)$





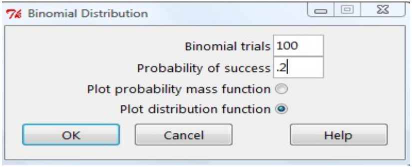
Plotting probability density function, pdf cont.

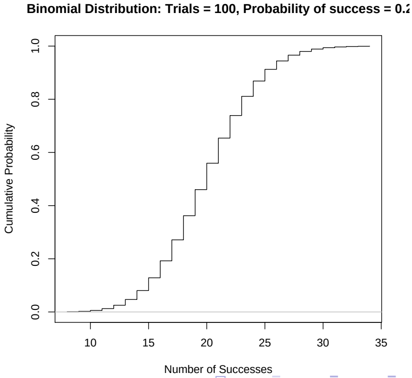
- For $X \sim Exp(\alpha = 4)$



Plotting (cumulative) distribution function, cdf

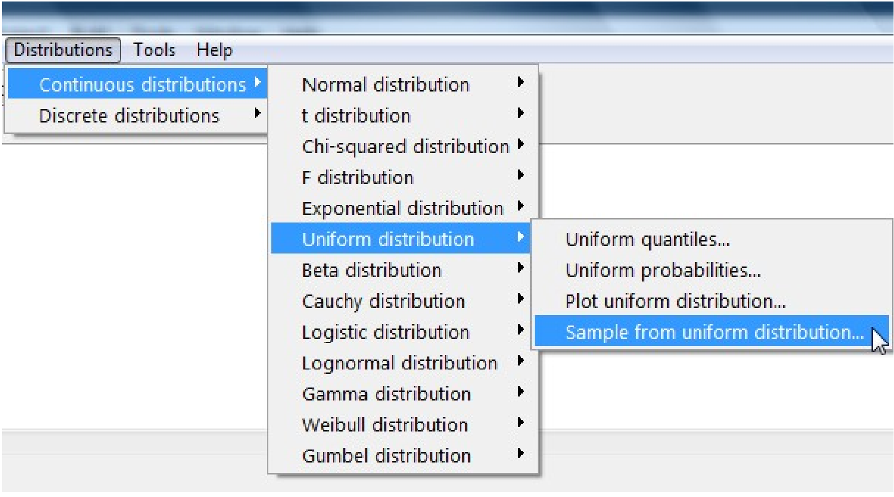
- For $X \sim Binomial(n = 100, p = 0.2)$





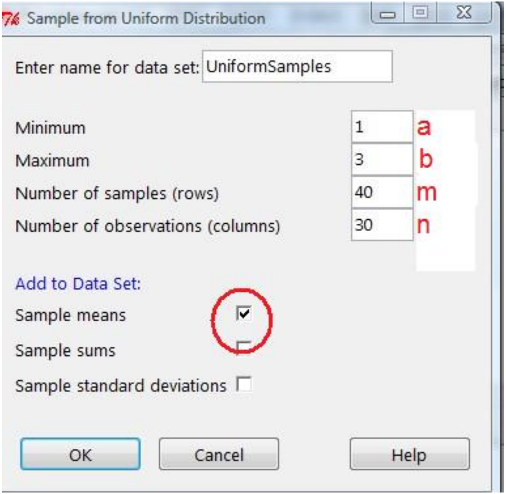
Generating samples from a distribution

- For $X \sim Uniform(a = 1, b = 3)$ generate $m = 40$ samples, each of size $n = 30$
- Calculate sample mean for each sample



Generating samples from a distribution cont.

- For $X \sim Uniform(a = 1, b = 3)$ generate $m = 40$ samples, each of size $n = 30$
- Calculate sample mean for each sample



Generating samples from a distribution cont.

Interpretation of the data table:

Sample 1: $\ldots, x_{28}^{(1)} = 1.88, x_{29}^{(1)} = 1.47, x_{30}^{(1)} = 1.26, \bar{x}^{(1)} = 2.05$

$\ldots$

Sample 10: $\ldots, x_{28}^{(10)} = 1.91, x_{29}^{(10)} = 2.78, \ldots, x_{30}^{(10)} = 2.70, \bar{x}^{(10)} = 1.96$

$\ldots$

UniformSamples

	obs22	obs23	obs24	obs25	obs26	obs27	obs28	obs29	obs30	mean
sample1	2.582253	1.553872	1.721375	1.669064	1.770635	1.490753	1.881825	1.473431	1.256755	2.052176
sample2	2.204278	1.416973	2.473223	1.773321	2.552076	2.521339	2.114723	1.171831	1.666728	1.899827
sample3	1.669251	2.606270	1.751927	1.589064	2.030841	2.604819	1.943639	2.256295	1.477837	2.079086
sample4	2.907687	1.244568	1.617775	1.261370	1.418457	2.064660	2.781578	2.543390	2.905500	2.051441
sample5	2.389902	1.461783	1.102419	2.815090	2.642343	1.002430	1.525546	1.152975	1.216409	2.028403
sample6	2.458287	2.139461	2.855624	1.922593	1.025467	2.725611	2.162475	2.574379	1.994768	2.028033
sample7	2.828883	1.968302	1.225310	2.366593	2.820637	2.822298	1.463490	1.756133	2.142561	2.073798
sample8	1.781358	2.357416	2.817104	1.257100	1.999403	2.744865	2.707661	1.280759	1.458416	1.874091
sample9	1.267791	2.754763	2.069404	1.231848	2.000610	1.079638	2.569111	2.139157	2.612367	1.951578
sample10	1.333785	1.773379	1.289907	2.715519	2.543643	1.544733	1.911046	2.780196	2.695635	1.958274
sample11	2.927192	1.675120	2.328251	2.202508	1.883138	2.855632	1.750483	1.234351	2.831332	2.127475
sample12	1.441371	1.695623	1.476890	2.393616	1.612549	2.040803	2.721733	2.332777	2.096245	1.896277
sample13	1.295602	1.086920	2.405013	2.529621	1.429409	1.016261	1.300232	1.003421	1.312333	1.838134
sample14	1.834232	1.491226	1.773037	1.848101	2.086949	1.481470	1.389542	1.820624	2.554322	1.998283
sample15	1.246469	2.516357	1.889279	2.265037	1.003413	1.003731	1.019298	2.981070	2.607242	1.986145
sample16	1.967234	2.080339	1.700127	2.315298	1.882857	1.902648	1.989172	1.540364	1.482040	1.989567
sample17	2.279792	2.838627	2.757429	2.717959	1.981486	2.818517	2.947425	1.694706	1.140786	2.197983
sample18	1.417441	1.108798	2.888227	1.900799	1.290917	1.035056	1.772251	1.429612	1.234399	1.960232
sample19	2.630464	2.588595	1.566644	2.580179	2.244479	2.468222	1.050039	2.420473	2.456681	2.233124
sample20	1.238690	2.125256	2.327209	1.159793	2.202307	1.203764	1.225207	2.345689	1.664489	1.819743

The Central Limit Theorem

Let X_1, X_2, ..., X_n be a random sample from a population distribution with mean μ and finite standard deviation σ. Then the sampling distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

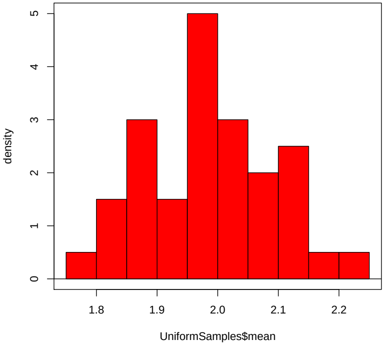
approaches a standard normal ditribution $N(0, 1)$ as $n \to \infty$.

- For highly skewed or heavy-tailed populations the samples may need to be larger for the distribution of the sample means shows a bell-shape.
- For any distribution (with finite standard deviation) the approximation tends to be better for larger sample sizes.

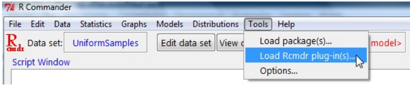
```
library(TeachingDemos)
example(clt.examp)
library(distrTeach)
illustrateCLT(Distr=Unif(), len=20)
```

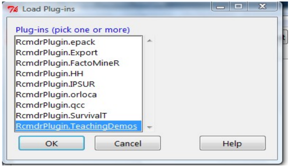
Towards Central Limit Theorem

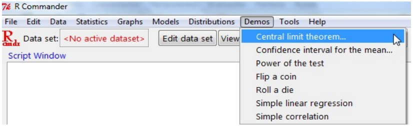
- Make a histogram of the $m = 40$ sample means from the previous page
- According to CLT, what should be its shape and its center? Normal, centered at the population mean $\mu = E[X] = \frac{a+b}{2} = 2$



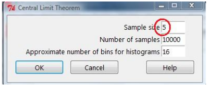
Central Limit Theorem with Teaching Demos

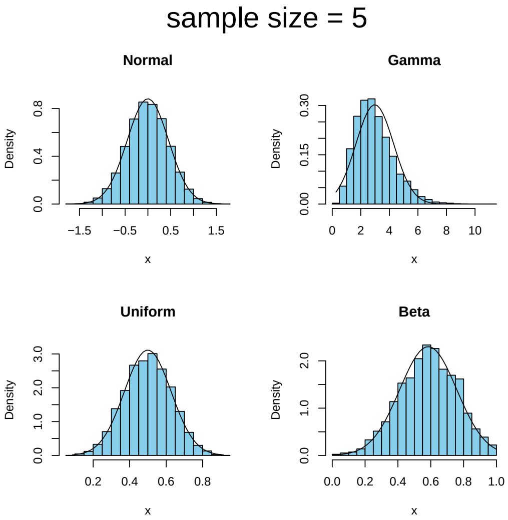






Central Limit Theorem with Teaching Demos, $n = 5$





Central Limit Theorem with Teaching Demos, $n = 35$