베이지안 1주차

Evloution of Statistical Thinking

- Frequentist vs Bayesian ## Historical Perspective
- late 17th/early 18 th centuries
- Limitation of Bayesian analyses:
    - Difficulty in evaluating $p(\theta|y)$ complex models analytically
    - Role of prior information : lack of objectivity
- Frequentist statistics was introduced as a way of overcoming these issues. 1920s mid-20th century
- Reemergence of Bayesian statistcs 1990's

Frequentist/ Classical Paradigm

- A parameter θ is viewed as an unknown fixed constant.
- Data are a repeatable random sample
- Goal : Estimate θ based on all available information (data) and find its associated error under asymptotic theory
- Inference is based on examining how well a procedure would do if it is used many times.:
    - Point estimates and standard errors or 95% confidence intervals
    - Deduction from $P(data|H_0)$, by setting α in advance.
    - Accept H_1 if $P(data|H_0) < \alpha$
    - Accept H_0 if $P(data|H_0) > \alpha$

Bayesian Paradigm

- A parameter θ is viewed as a random variable whose distribution is unknown, and descirbed probabilistically.
- Data are observed from the realized sample
- Goal : Estimate the distribution of θ conditional on the observed adata, the posterior distribution of θ.
- Inference is based on summaries of the posterior distribution of θ:
    - Induction from $P(\theta|data)$, starting with $P(\theta)$
    - Broad descriptions of the posterior distribution such as means and quantiles.
    - Highest posterior density intervals indicating region of highgest posteriror probabilty regardless of contiguity.

Differences Between Frequentist and Bayesian

- What is fixed?
    - Frequentist : Parameters are fixed
    - Bayesian : Data are fixed
- General inference:
    - Frequentist : $P(data|\theta)$ is sampling distribution of the data given the parameter.
    - Bayesian : $P(\theta)$ is the prior distribution of the parameter (before the data are seen) and $P(\theta|data)$ is the posterior distribution of the parameter.
- 95% Intervals:
    - Frequentist : In repeated samples, 95% of realized intervals covers the true parameter.
    - Bayesian : For the se data, with probability 95% the parameter is in the interval.
- Bayesian inference proceeds vertically, with x fixed, according to the posterior distribution $g(\mu|x)$.
- Frequentists reason horizeontally, with μ fixed and x varying.

Overall Recommendation

- Be pragmatic, not dogmatic:
    - Use what has been shown to work.
    - As a default approach, the following will serve you well: Design as a Bayesian, and evaluate as a frequentist.
- Construct models and procedures from a Bayesian perspective, and use frequentist tools to evaluate their empirical and theoretical performance.
- In the spirit of being pragmatic, it might seem unnecessarily restrictive to limit oneself to Bayesian procedures, and indeed, there are times when a non-Bayesian procedure may be preferable to a Bayesian one.
- However, typically, it turns out that there is no disadvantage in considering only Bayesian procedures.

Probabilities Defined on Events

- Consider an experiment whose sample space is S
- For each event A of the sample space S, we assume that a number $P(A)$ is defined and satisfies the following three conditions:
    - $0 \leq P(A) \leq 1$
    - $P(S) = 1$
    - For any sequence of events $A_1, A_2, ...$ that are pairwise mutually exclusive, that is events for which $A_n \cap A_m = \phi$, when $n \neq m$, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ We refer to $P(A)$ as the probability of the event A

Conditional Probabilities

- If the event B occurs, then in order for A to occur it is necessary for the actual occurence to be a point in both A and in B, that is, it must be in $A \cap B$. Now, because we know that B has occured, it follows that B becomes our new sample space and hence the probability that the event $A \cap B$ occurs will equal the probability of $A \cap B$ relative to probability of B. That is. $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Independent Events

- Two events A and B are said to be independent: $P(A \cap B) = P(A)P(B)$ which implies that $P(A|B) = P(A)$ $P(B|A) = P(B)$
- More generally, the events $A_1, ..., A_n$ are said to be independent if for every subset $A'_1, ..., A'_n$ $(r \leq n)$ of these event: $P(A'_1, ..., A'_r) = P(A'_1)...P(A'_r)$

Law of Total Probability

If events $A_1, ..., A_k$ partition a sample space S into mutually exclusive and exhaustive nonempty events, then the Law of Total Probability states that the total probability of an event B is given by: $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_k \cap B)$
$= P(B|A_1)P(A) + ... + P(B|A_k)P(A_k) = \sum_{j=1}^{k} P(B|A_j)P(A_j)$

Bayes' Theorem

- Bayes' Theorem provides a method for inverting conditional probabilities. In its simplest form, if A and B are events and $P(B) > 0$, then $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B^c|A)P(A)}$

Distribution Review

Bernoulli Trial

- Several discrete distribution can formulated in terms of the outcomes of Bernoulli trials
- A Bernoulli trial has exactly two possible outcomes, "success" or "failure"

- A Bernoulli random variable X has the probability mass function $P(X = 1) = p$ and $P(X = 0) = 1 - p$ where p is the probability of success, $0 \leq p \leq 1$.
- $X \sim Bern(p)$

R code

- Rlab package
- pdf

```
library(Rlab)
```

```
## Rlab 2.15.1 attached.
```

```
##
## Attaching package: 'Rlab'
```

```
## The following objects are masked from 'package:stats':
##
##      dexp, dgamma, dweibull, pexp, pgamma, pweibull, qexp, qgamma,
##      qweibull, rexp, rgamma, rweibull
```

```
## The following object is masked from 'package:datasets':
##
##      precip
```

```
x_dbern <- seq(0, 10, by = 1)
y_dbern <- dbern(x_dbern, prob=0.7)
plot(y_dbern, type = "o")
```
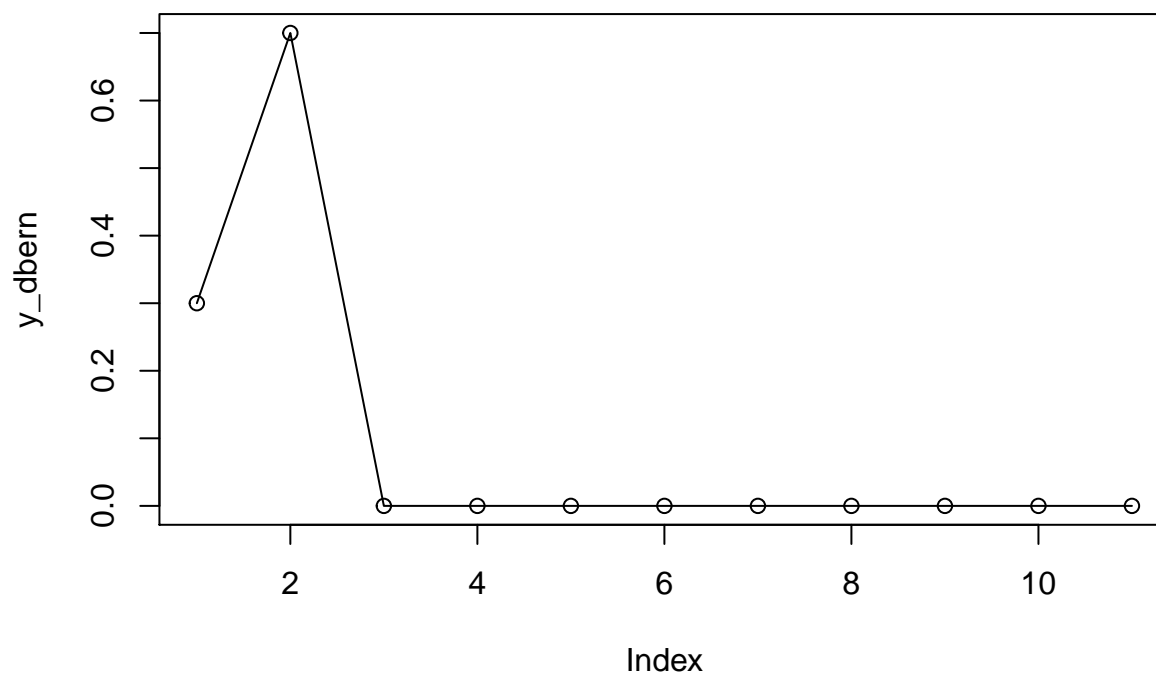

* cdf

```
x_pbern <- seq(0, 10, by=1)
y_pbern <- pbern(x_pbern, prob = 0.7)
plot(y_pbern, type="o")
```
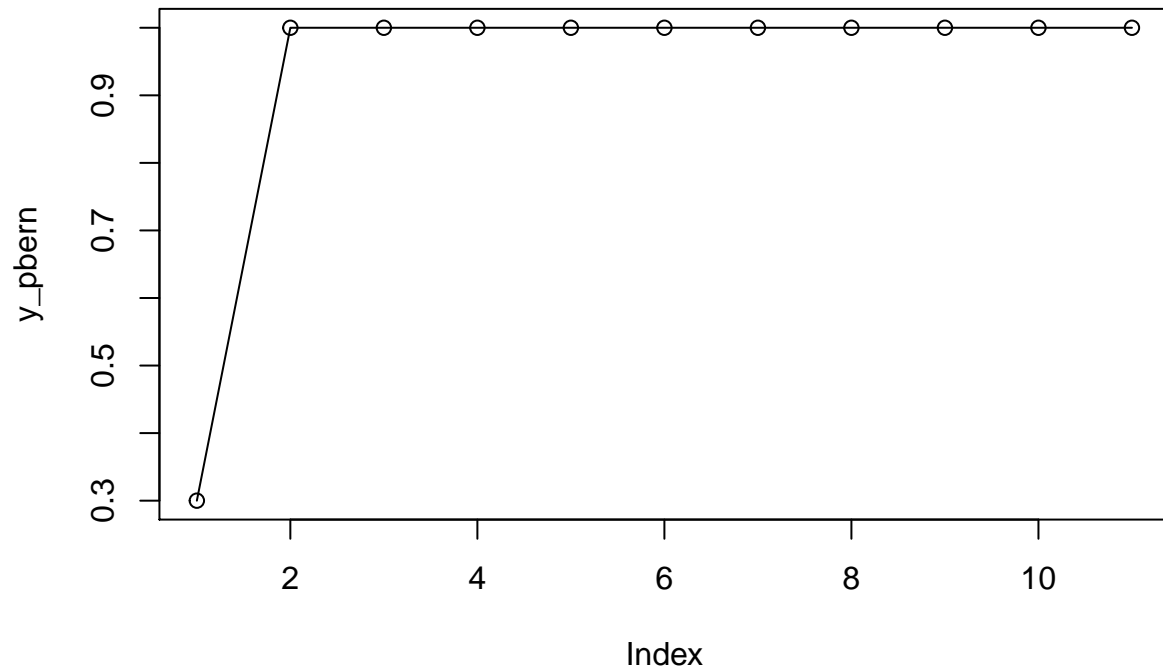

* Quantile Function

```
x_qbern <- seq(0, 1, by = 0.1)
y_qbern <- qbern(x_qbern, prob = 0.7)
plot(y_qbern, type="o")
```
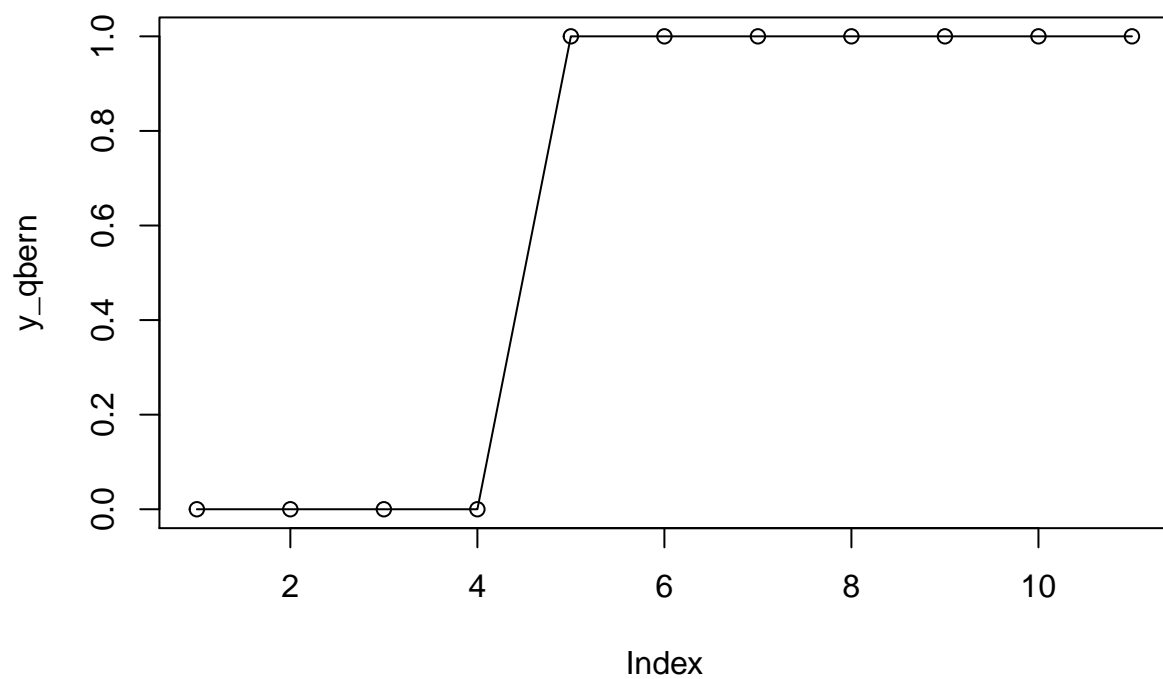
* Generating Random Numbers

```
set.seed(20211113)
N <- 10000
y_rbern <- rbern(N, prob = 0.7)
hist(y_rbern, breaks = 5, main = "")
```
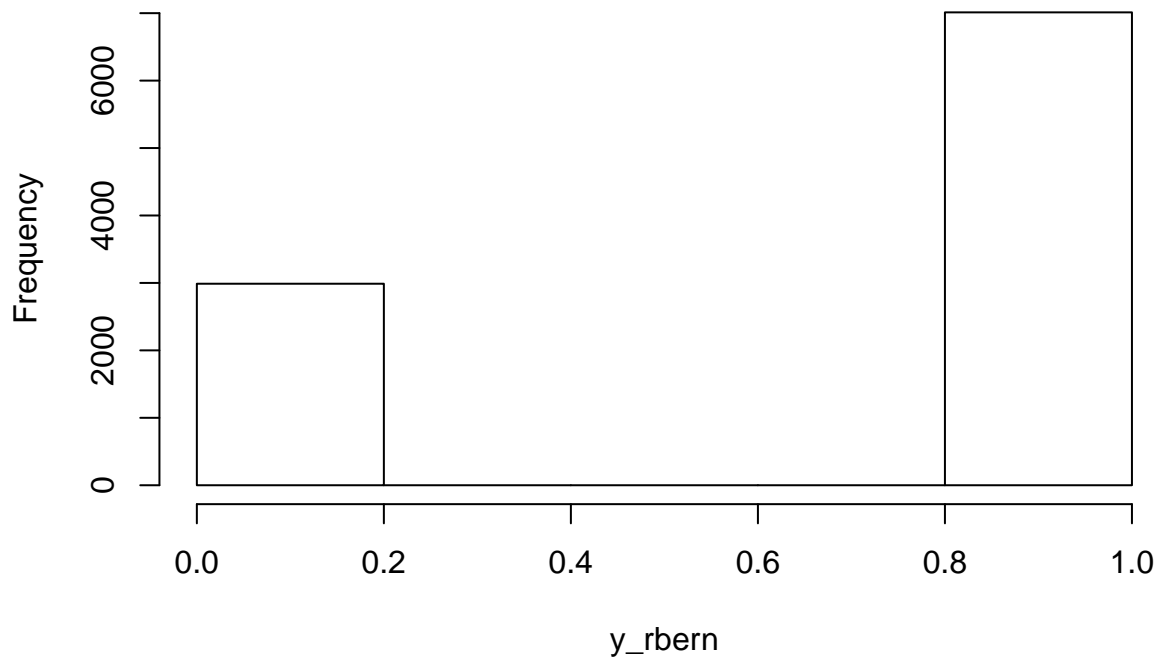
Binomial Distribution

- $X_1, ..., X_m \overset{iid}{\sim} Bern(p)$ => $X = \sum_{i=1}^{n} X_i \sim Bin(n, p)$
- Suppose that X records the number of success in n iid Bernoulli trials with success probability p.
- Then X has the Binomial(n, p) distribution with $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
  $= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ $x = 0, 1, ..., n$
- The mean and variance are $E[X] = np$ and $Var(X) = np(1-p)$ #### R code
- pdf

```
library(stats)
n <- 60
x_dbinom <- seq(0, n, by = 1)
y_dbinom <- dbinom(x_dbinom, n, 0.7)
plot(x_dbinom, y_dbinom)
```
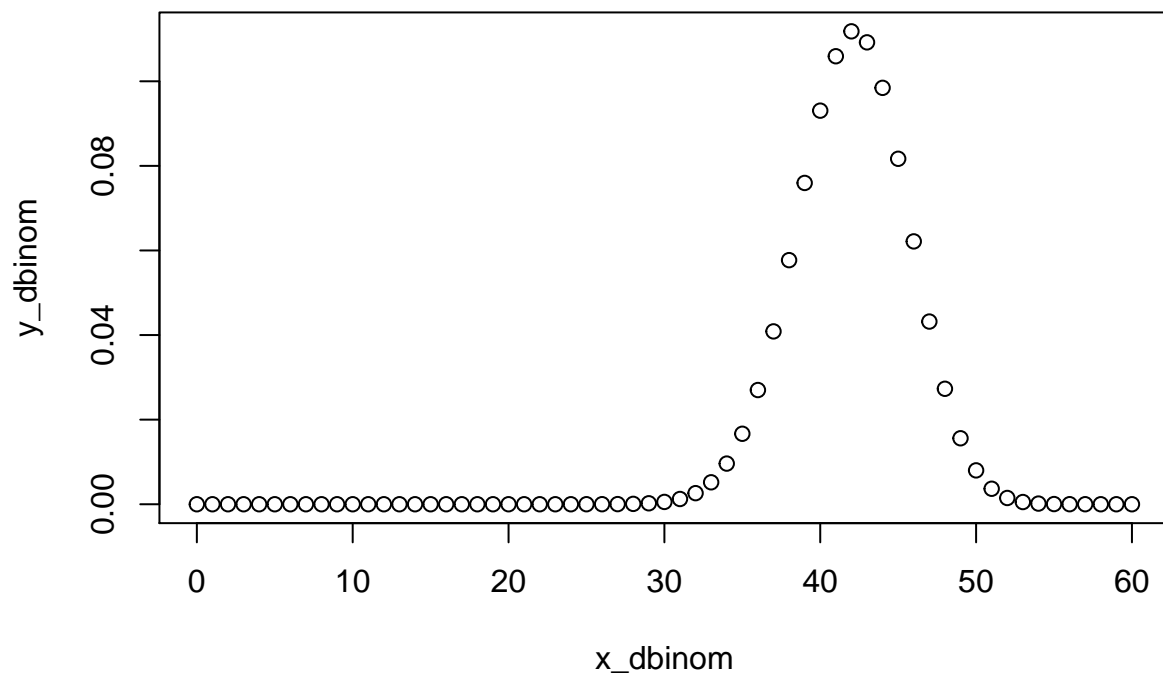
* cdf

```
n <- 60
x_pbinom <- seq(0, n, by=1)
y_pbinom <- pbinom(x_pbinom, size = n, prob = 0.7)
plot(y_pbinom, type="o")
```
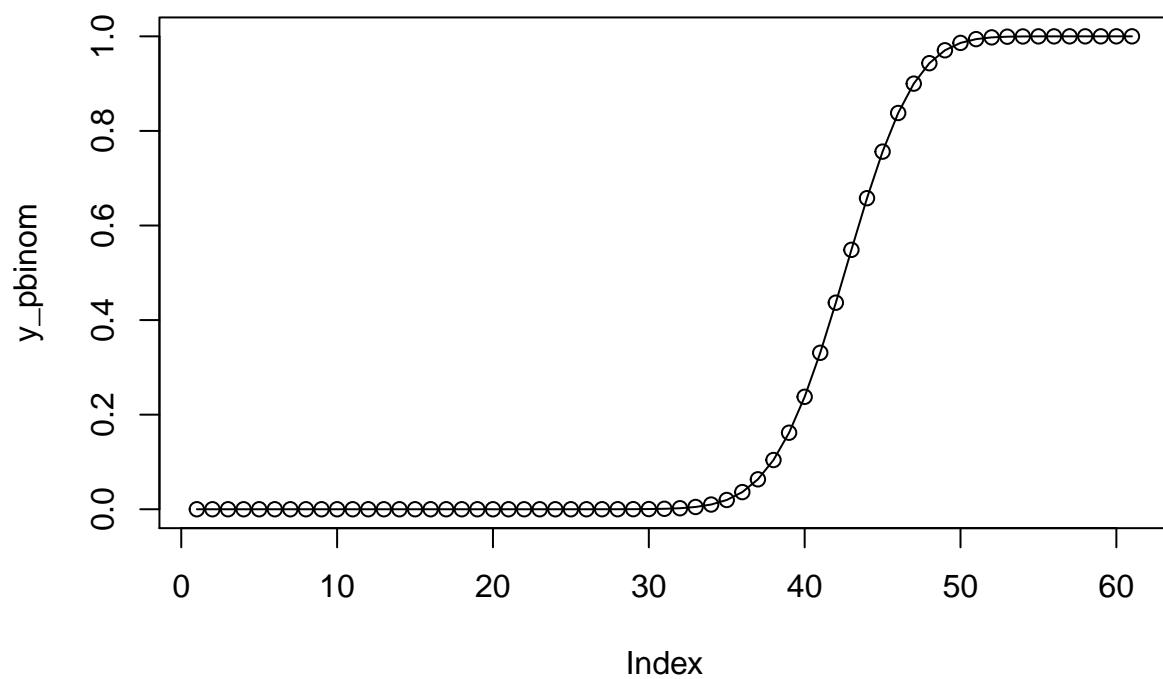
* Quantile Function

```
n <- 60
x_qbinom <- seq(0, 1, by = 0.1)
y_qbinom <- qbinom(x_qbinom, size = n, prob = 0.7)
plot(y_qbinom, type="o")
```
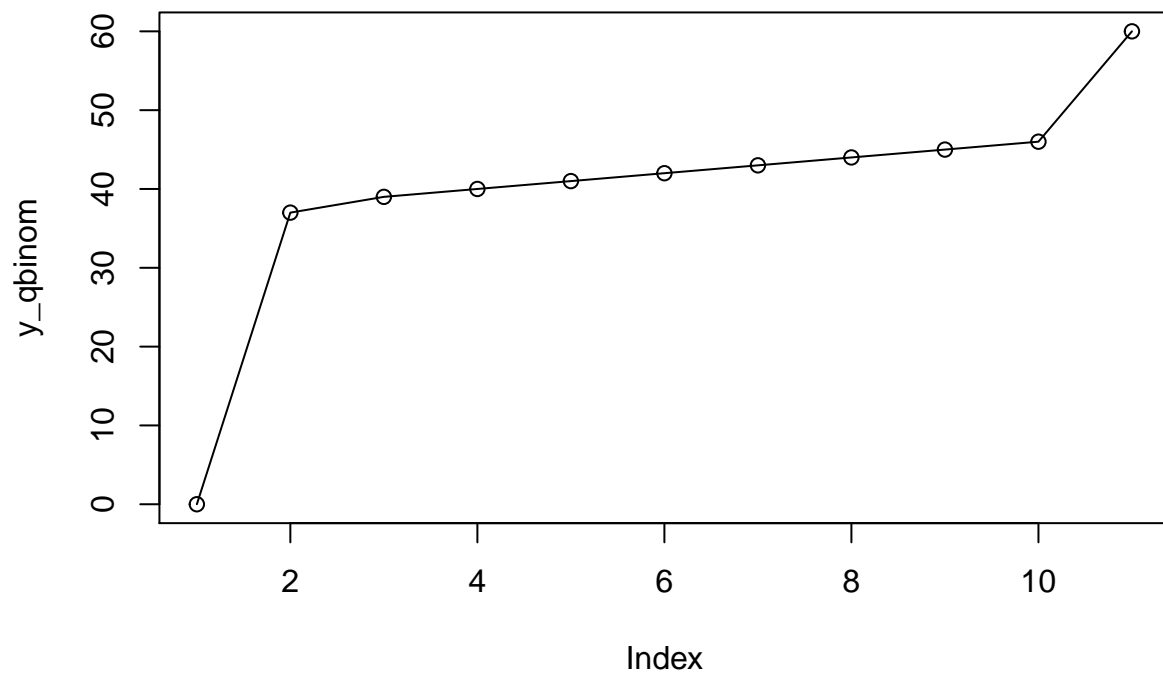
* Generating Random Numbers

```r
set.seed(20211113)
N <- 100
y_rbinom <- rbinom(N, size= N, prob = 0.7)
hist(y_rbinom, breaks = 5, main = "")
```
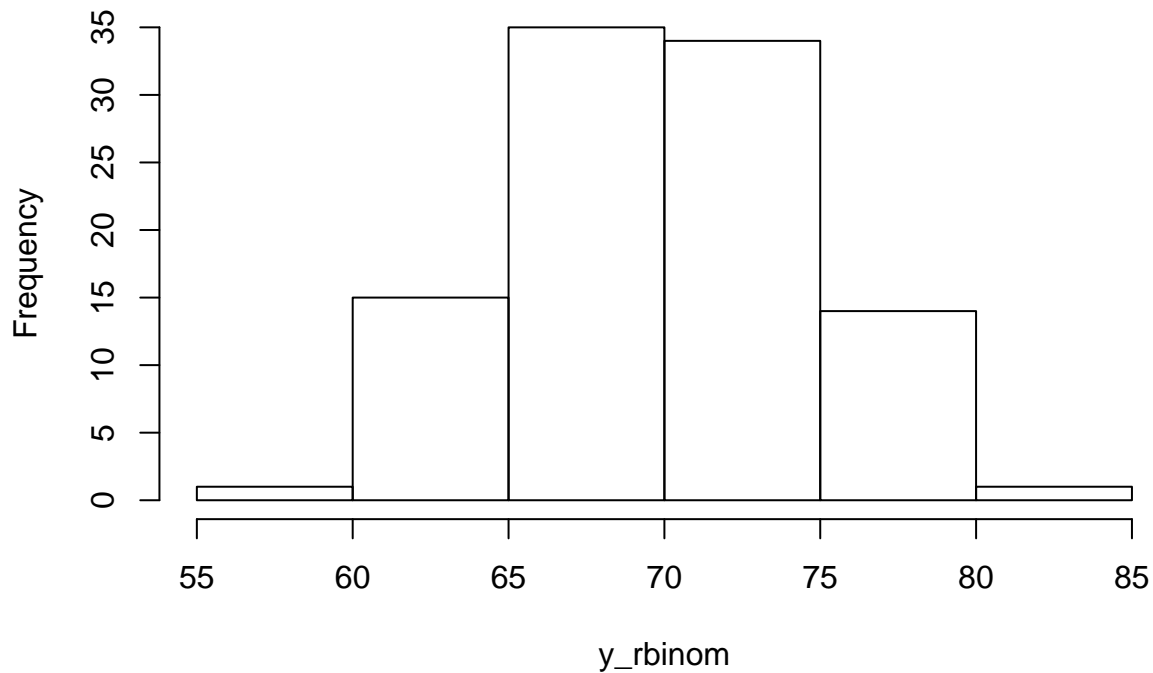
Multinomial Distribution

- A multinomial distribution is a generalization of the binomial distribution
- Suppose one does an experiment of extracting n balls of k different colors from a bag, replacing the extracted ball after each draw. Balls from the same color are $(i = 1, ..., k)$ as X_i and devote as p_i the probability that given extraction will be in color i.
- Then $X_1, ..., X_k$ has the multinomial distribution iwth joint pdf: $f(x_1, ..., x_p) = P(X_1 = x_1, ..., X_k = x_k)$ $= \frac{n!}{x_1!...x_n!}p_1^{x_1}...p_k^{x_k}I(\sum_{i=1}^{k} x_i = n)$
- If follows that: $E[X_i] = np_i, \ Var(X_i) = np_i(1 - p_i), \ Cov(X_i, X_j) = -np_ip_j$ for $i \neq j$

R code

- pdf

```
N <- 12
x_dmultinom <- c(2, 3, 7)
dmultinom(x_dmultinom, size=N, prob=c(0.1, 0.2, 0.7))
```

```
## [1] 0.05217968
```

- Generating Random Numbers

```
N <- 12
k <- 10
rmultinom(k, size = N, prob = c(0.1,0.2,0.7))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    1    1    3    3    1    0    2    2    2    2
```

```
## [2,]    3    5    3    1    2    2    3    2    2    2
## [3,]    8    6    6    8    9   10    7    8    8    8
```

Geometric Distribution

- Suppose that independent Bernoulli trials, each having probability p of being a success, are performed, until a success occurs. If we let X be the number of trials required until the first success, then X has the geometric distribution with pdf $P(X = x) = (1-p)^{x-1}p, x = 1, 2, \ldots$
- The cdf of X is $F(x) = P(x \le x) = 1 - (1-p)^x$
- It follows that $E[X] = \frac{1}{p}$ and $Var(x) = \frac{1-p}{p^2}$

R code

- pdf

```
library(Rlab)
x_dgeom <- seq(0, 10, by = 1)
y_dgeom <- dgeom(x_dgeom, prob=0.2)
plot(y_dgeom, type = "o")
```
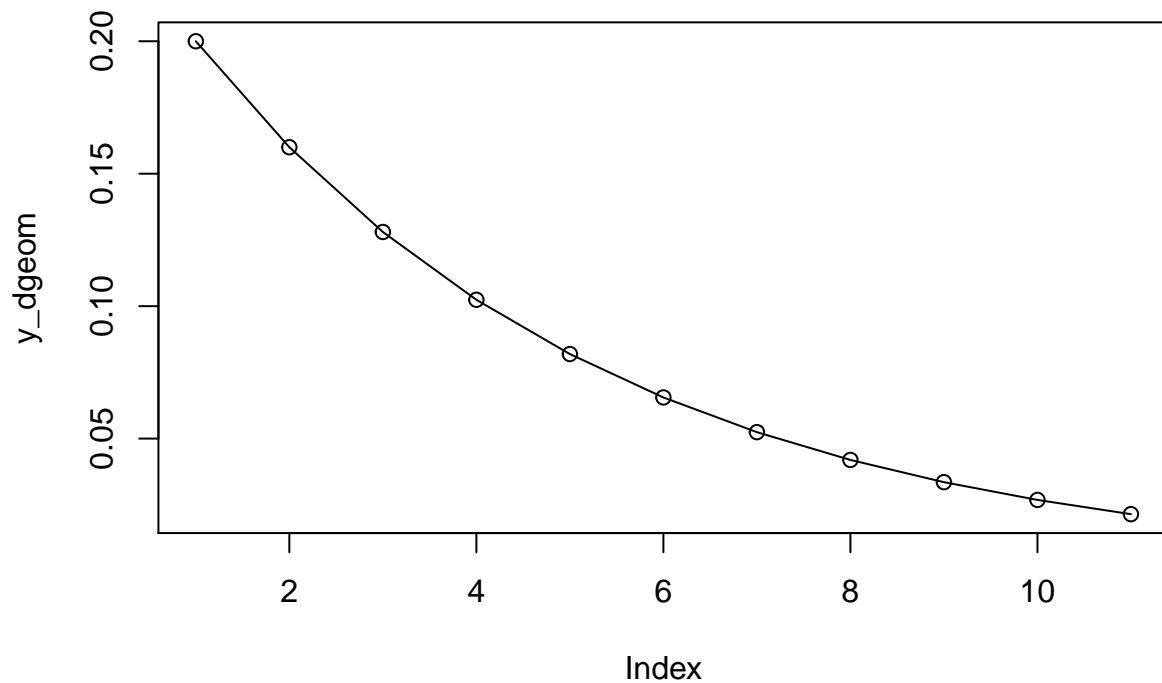

* cdf

```
x_pgeom <- seq(0, 10, by=1)
y_pgeom <- pgeom(x_pgeom, prob = 0.2)
plot(y_pgeom, type="o")
```
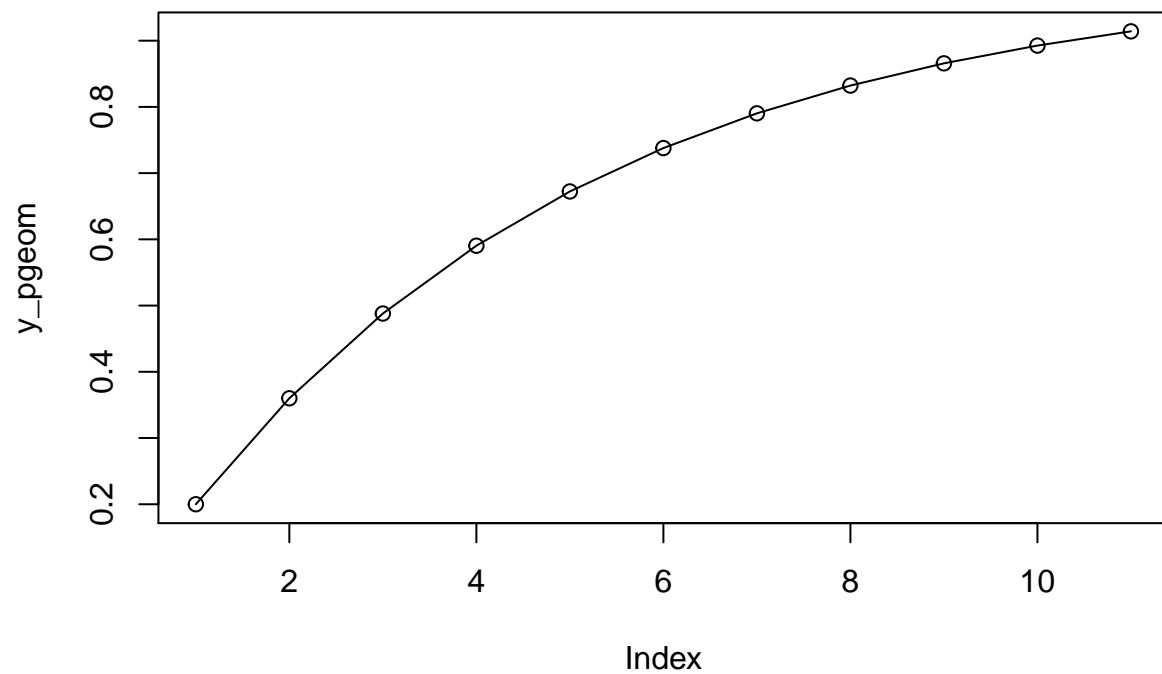
* Quantile Function

```
x_qgeom <- seq(0, 1, by = 0.1)
y_qgeom <- qgeom(x_qgeom, prob = 0.2)
plot(y_qgeom, type="o")
```
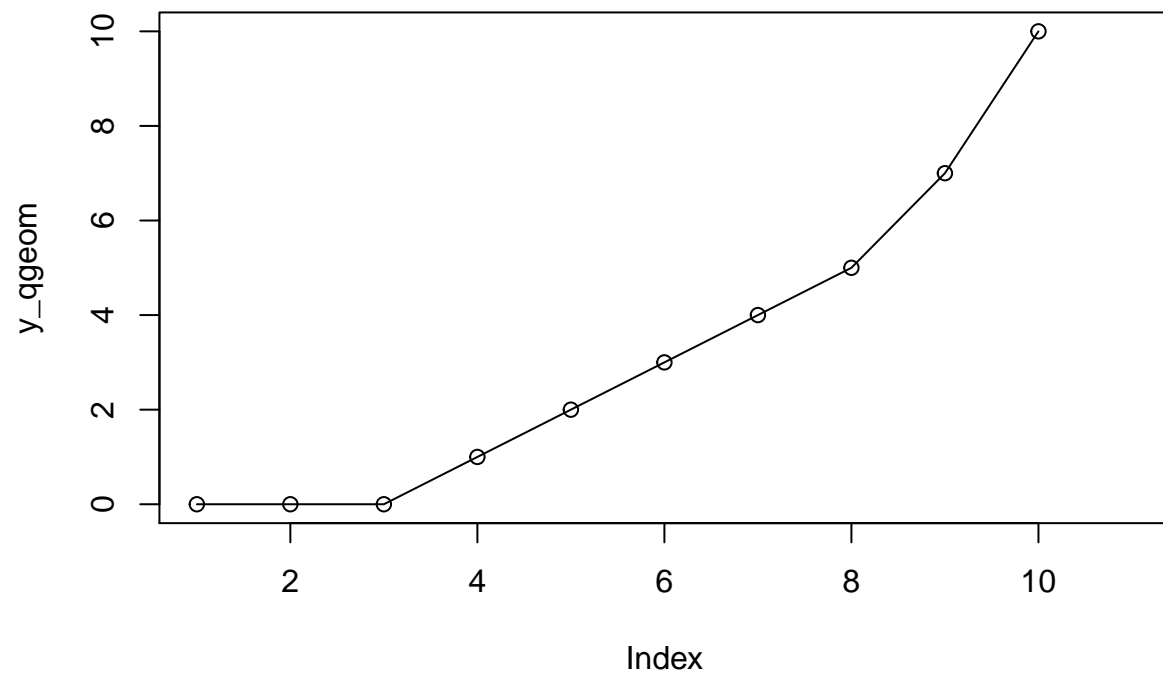
* Generating Random Numbers

```r
set.seed(20211113)
N <- 10000
y_rgeom <- rgeom(N, prob = 0.2)
hist(y_rgeom, breaks = 100, main = "")
```
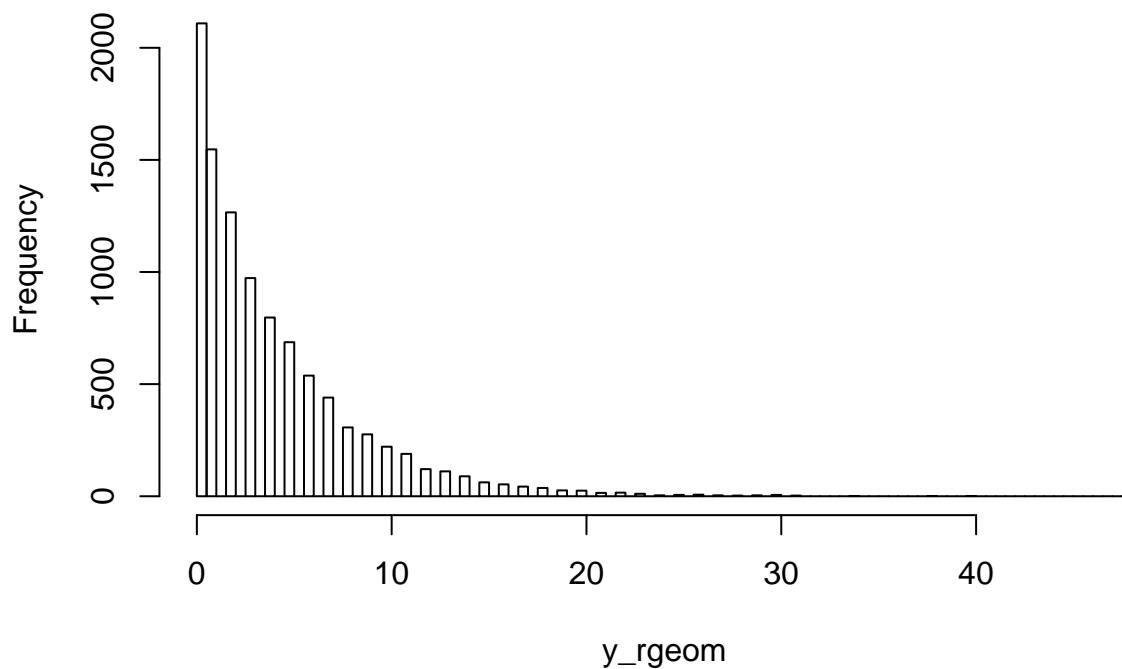
Negative Binomial Distribution (generalization of geometric distribution)

- The distribution that applies to the random variable X equal to the number of the trial on which the r th success occurs $(r = 2, 3, 4, ...)$ is the negative binomial distribution with pdf: $P(X = x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}, x = r, r+1, r+2, ...$
- The geometric distribution is a special case of the negative binomial distribution with $r = 1$.
- It follows that: $E[X] = \frac{r}{p}$ and $Var(X) = \frac{r(1-p)}{p^2}$

R code

- pdf

```
N <- 12
x_dnbinom <- seq(0, 10, by = 1)
y_dnbinom <- dnbinom(x_dnbinom, size=N, prob=0.2)
plot(y_dnbinom, type = "o")
```

* cdf
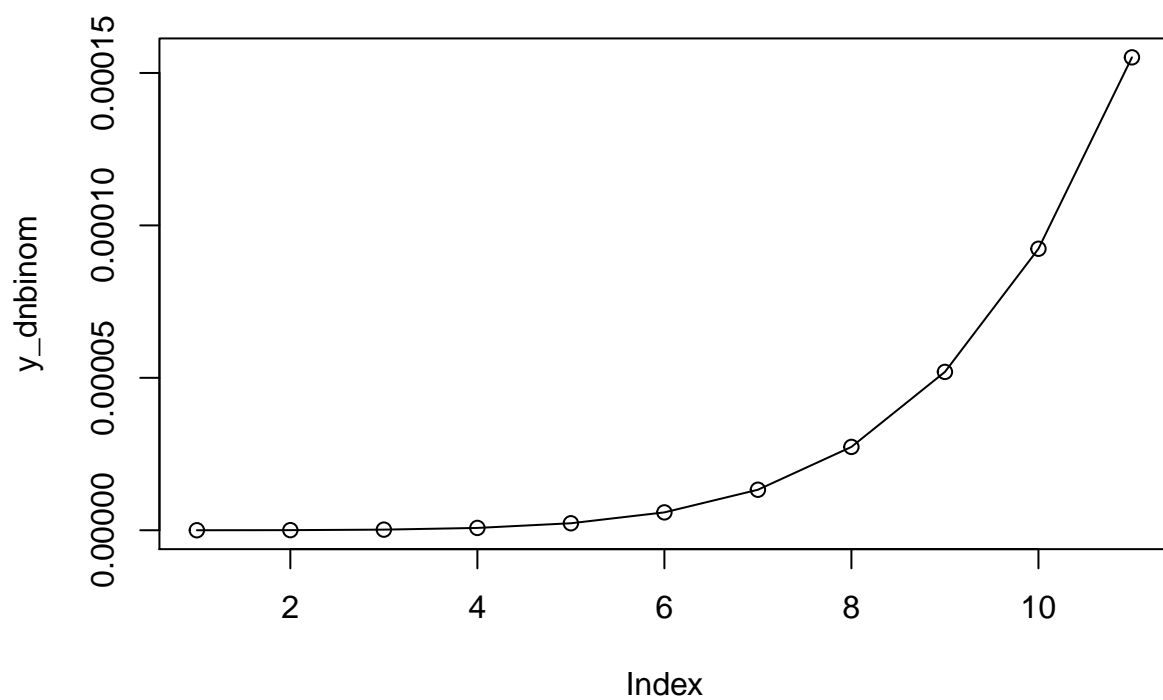
```r
N <- 12
x_pnbinom <- seq(0, 10, by=1)
y_pnbinom <- pgeom(x_pnbinom, prob = 0.2)
plot(y_pnbinom, type="o")
```
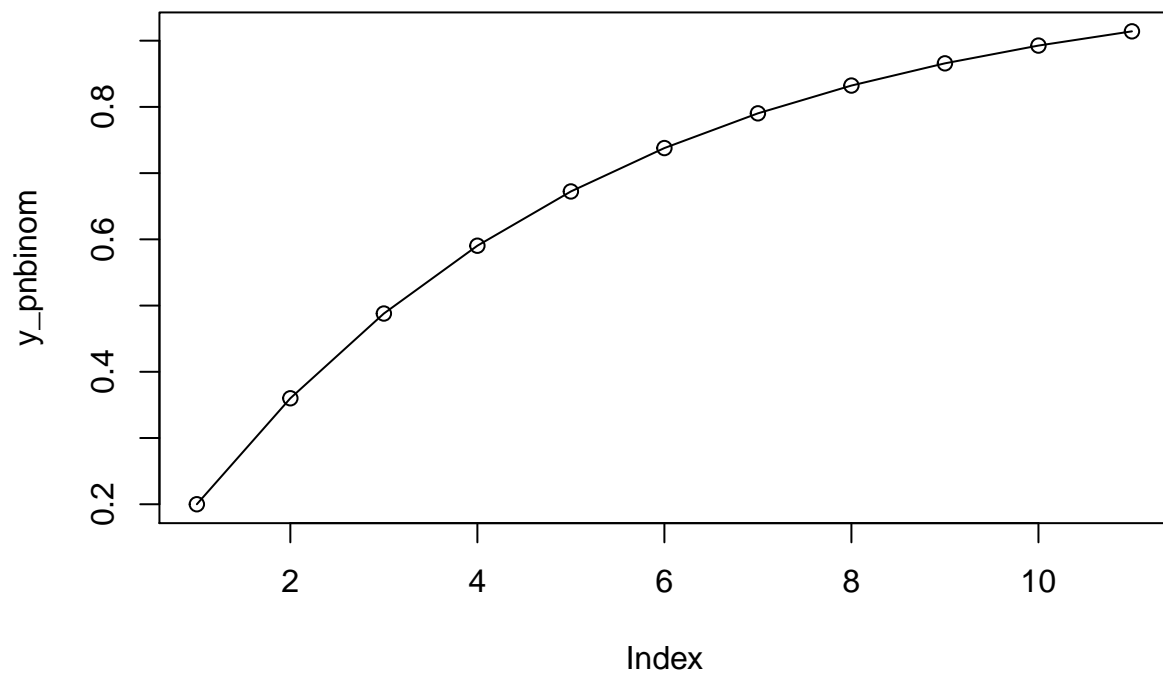
* Quantile Function

```
N <- 12
x_qnbinom <- seq(0, 1, by = 0.1)
y_qnbinom <- qnbinom(x_qnbinom, size=N, prob = 0.2)
plot(y_qnbinom, type="o")
```
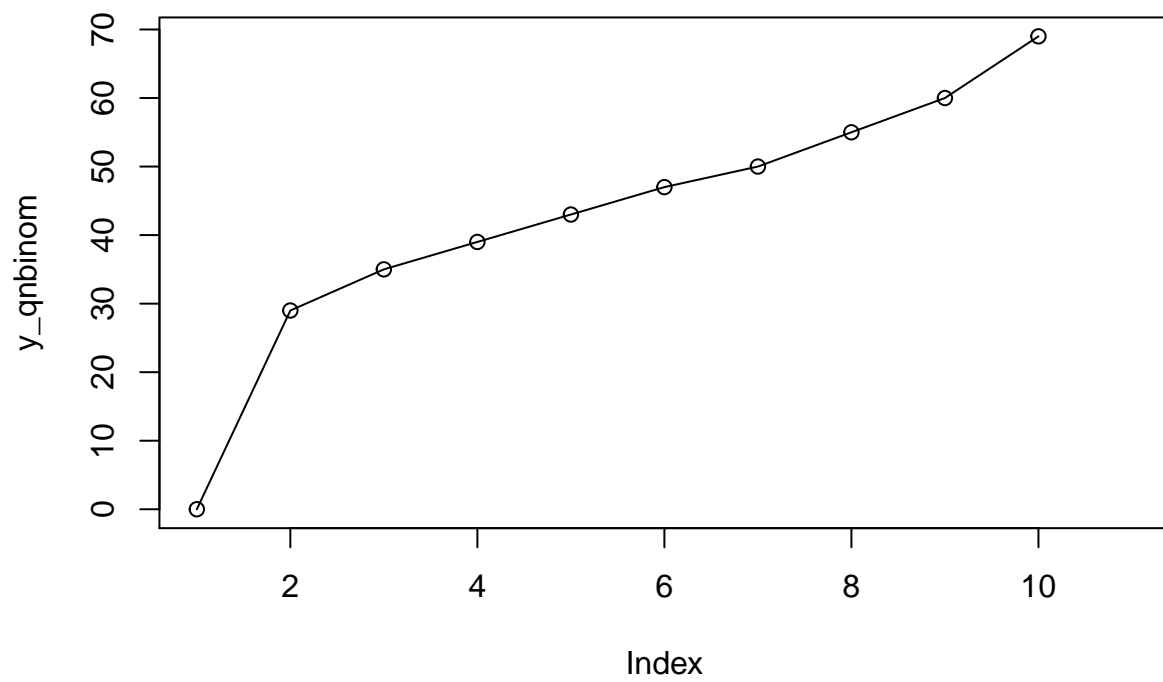
* Generating Random Numbers

```
set.seed(20211113)
N <- 1000
y_rnbinom <- rbinom(N, size=N, prob = 0.2)
hist(y_rnbinom, breaks = 100, main = "")
```

Frequency (y-axis), **y_rnbinom** (x-axis)
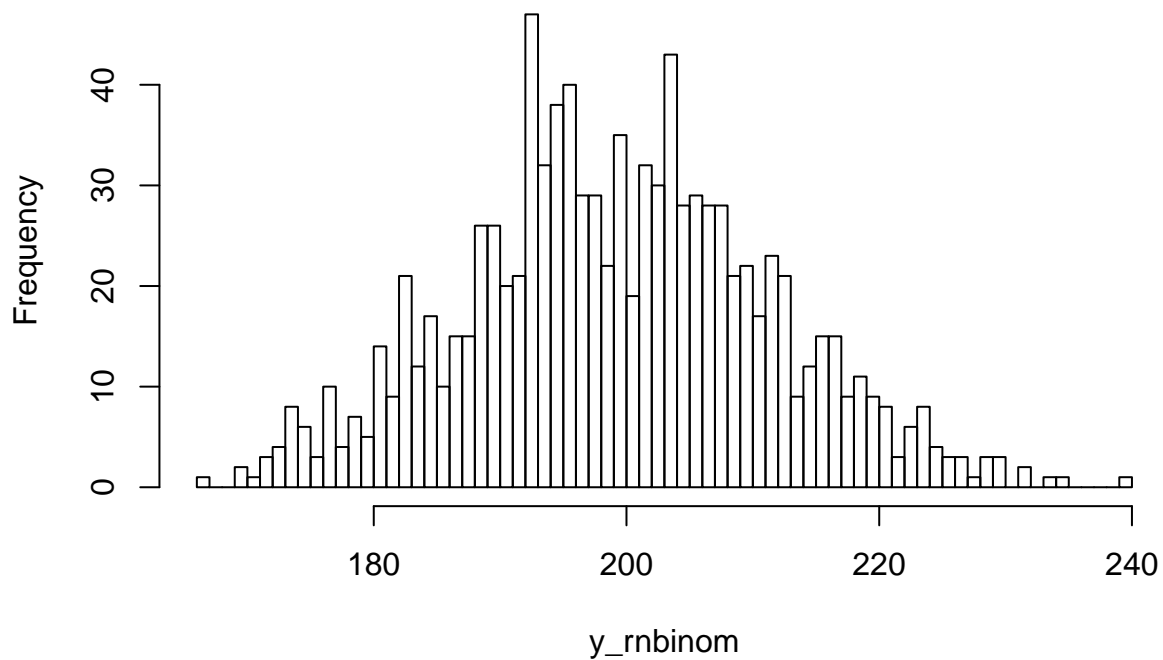
Examples

- A geological study indicates that an explantory oil well drilled in a particular region should be strike oil with probability 0.2. Find the probability that the third oil strike comees on the fifth well drained.

Poisson Distribution

- The Poisson distribution is ad discrete probability distribution that expresses the probability of a given number X of events occuring in a find interval of time and/or space with pdf: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ...$ where $\lambda > 0$ is the average value of X
- It follows that: $E[X] = \lambda$ and $Var(X) = \lambda$

R code

- pdf

```
x_dpois <- seq(0, 10, by = 1)
y_dpois <- dpois(x_dpois, lambda=5)
plot(y_dpois, type = "o")
```
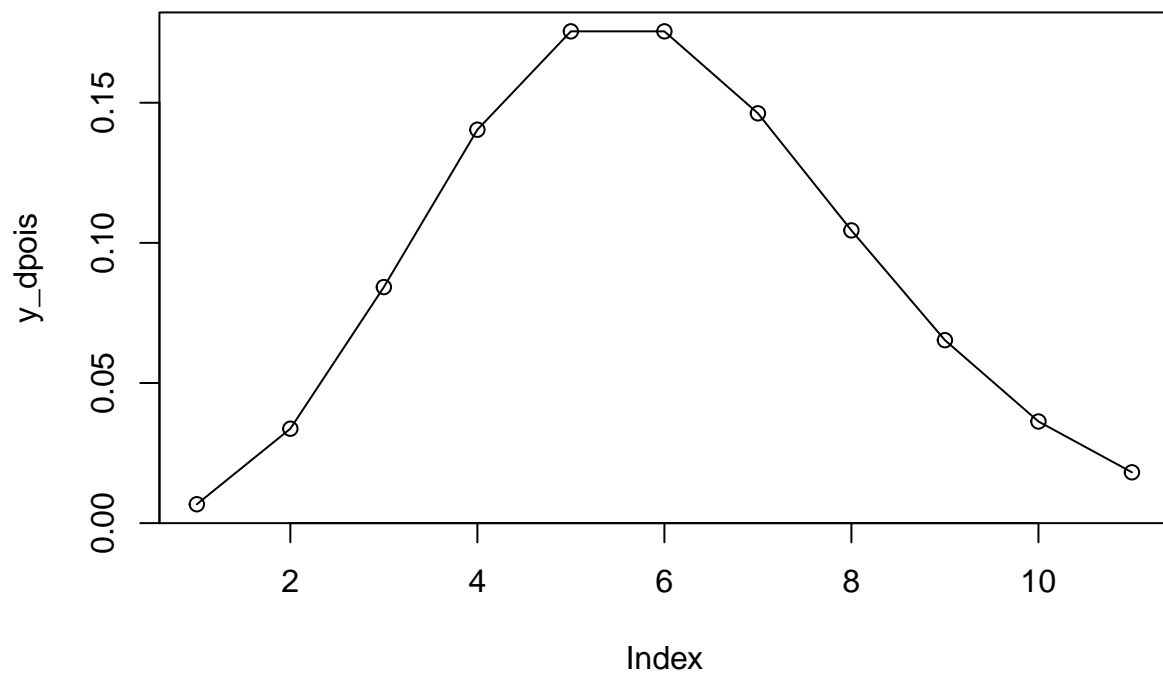
* cdf

```
x_ppois <- seq(0, 10, by=1)
y_ppois <- ppois(x_ppois, lambda=5)
plot(y_ppois, type="o")
```
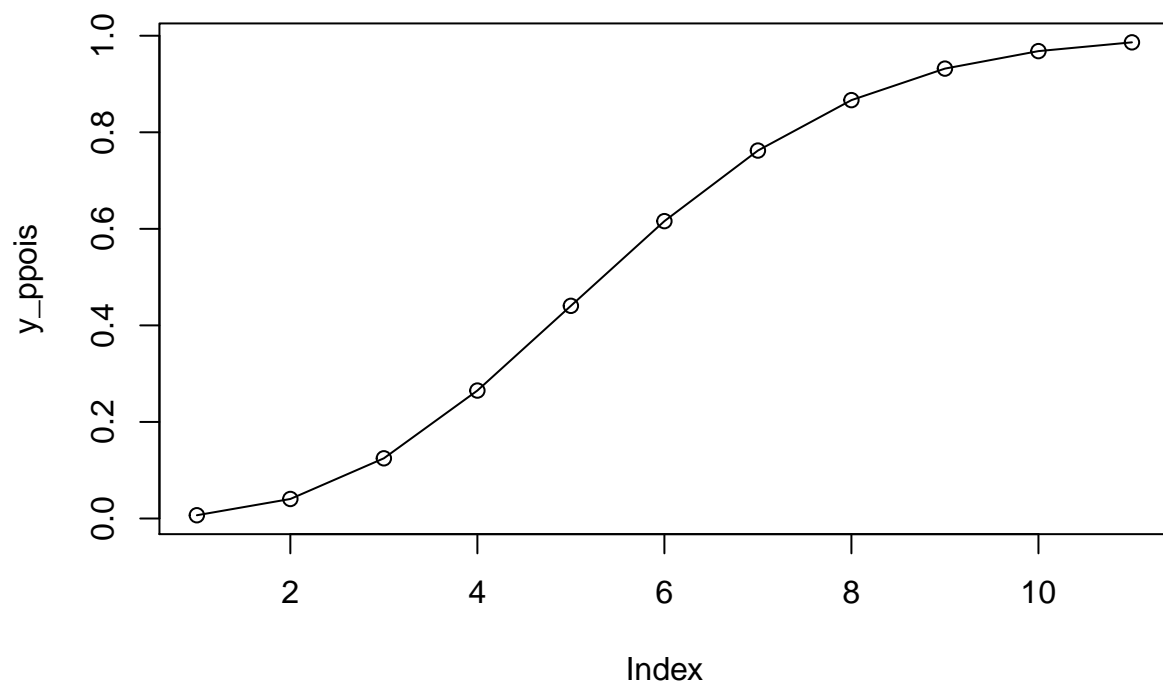
* Quantile Function

```
x_qpois <- seq(0, 1, by = 0.1)
y_qpois <- qpois(x_qpois, lambda=5)
plot(y_qpois, type="o")
```
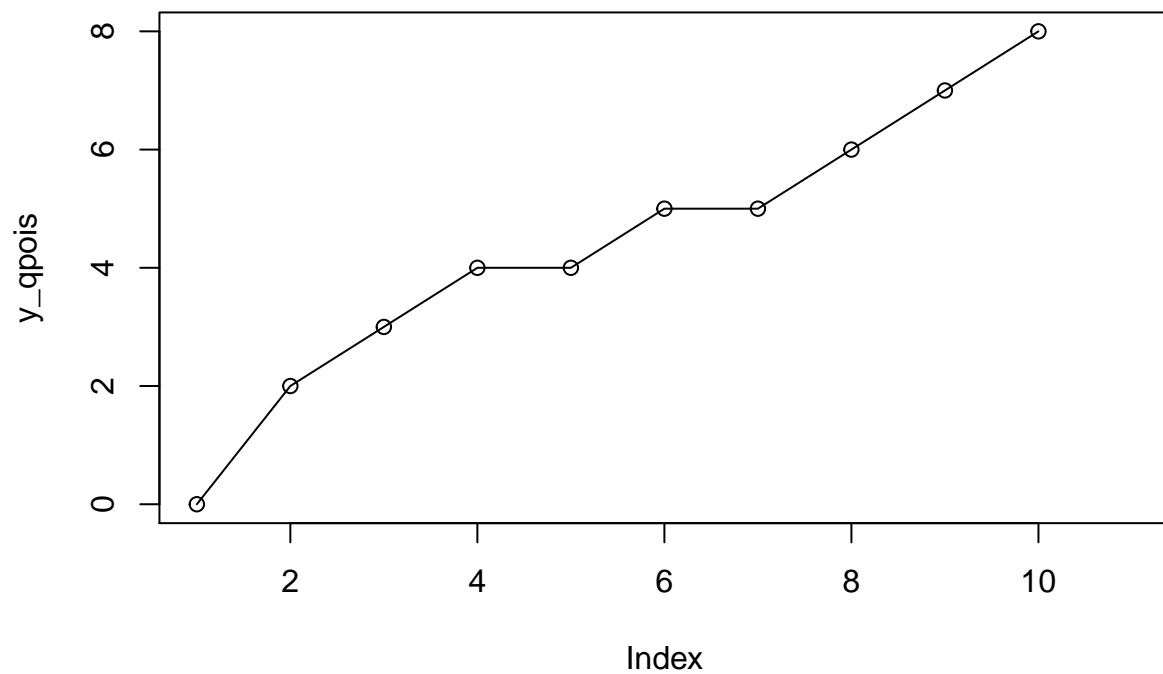
* Generating Random Numbers

```r
set.seed(20211113)
N <- 1000
y_rpois <- rpois(N, lambda = 5)
hist(y_rpois, breaks = 100, main = "")
```
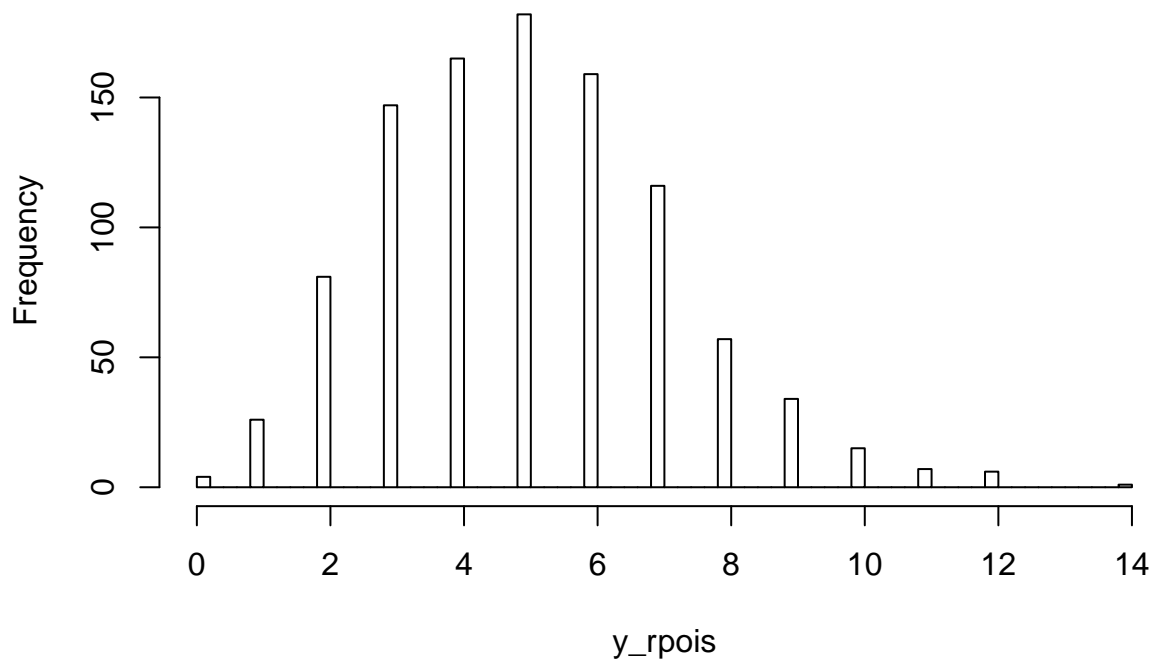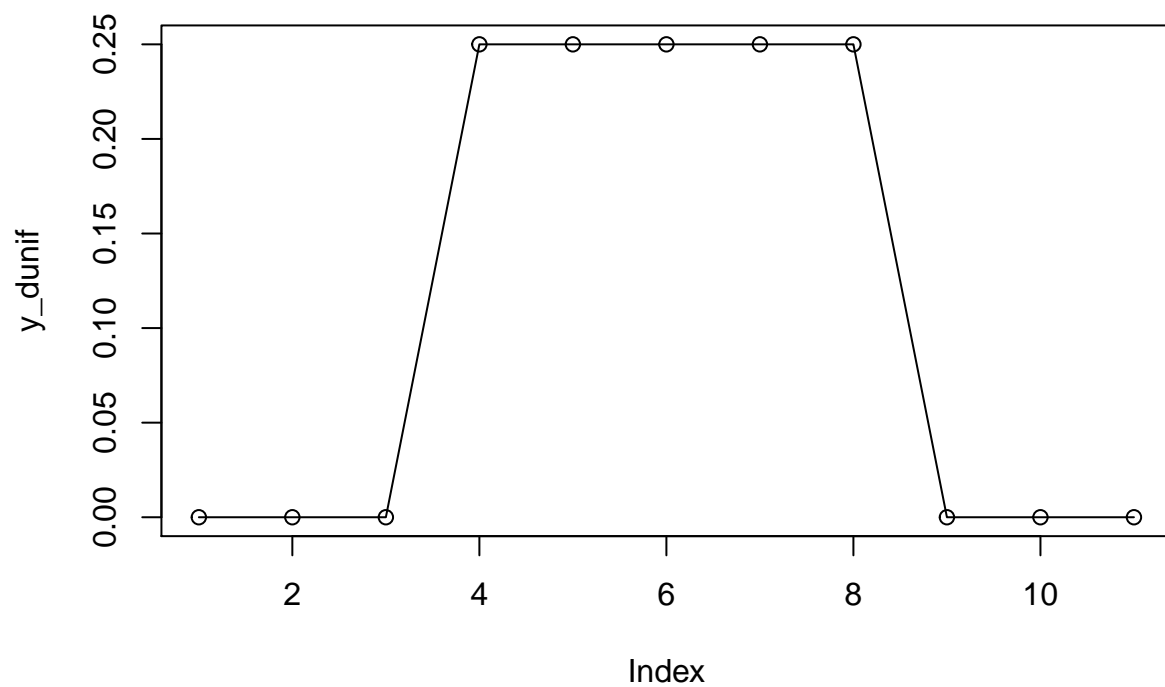
Uniform Distribution

- It $\theta_1 < \theta_2$, a random variable X is said to have a continuous uniform distribution on the interval (θ_1, θ_2) with pdf: $f(x) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 \leq x \leq \theta_2)$
- It follows that: $E[X] = \frac{\theta_1 + \theta_2}{2}$ and $Var(X) = \frac{(\theta_2 - \theta_1)^2}{12}$

R code

- pdf

```
x_dunif <- seq(0, 10, by = 1)
y_dunif <- dunif(x_dunif, min=3, max=7)
plot(y_dunif, type = "o")
```

* cdf

```
x_punif <- seq(0, 10, by=1)
y_punif <- punif(x_punif, min=2, max=7)
plot(y_punif, type="o")
```
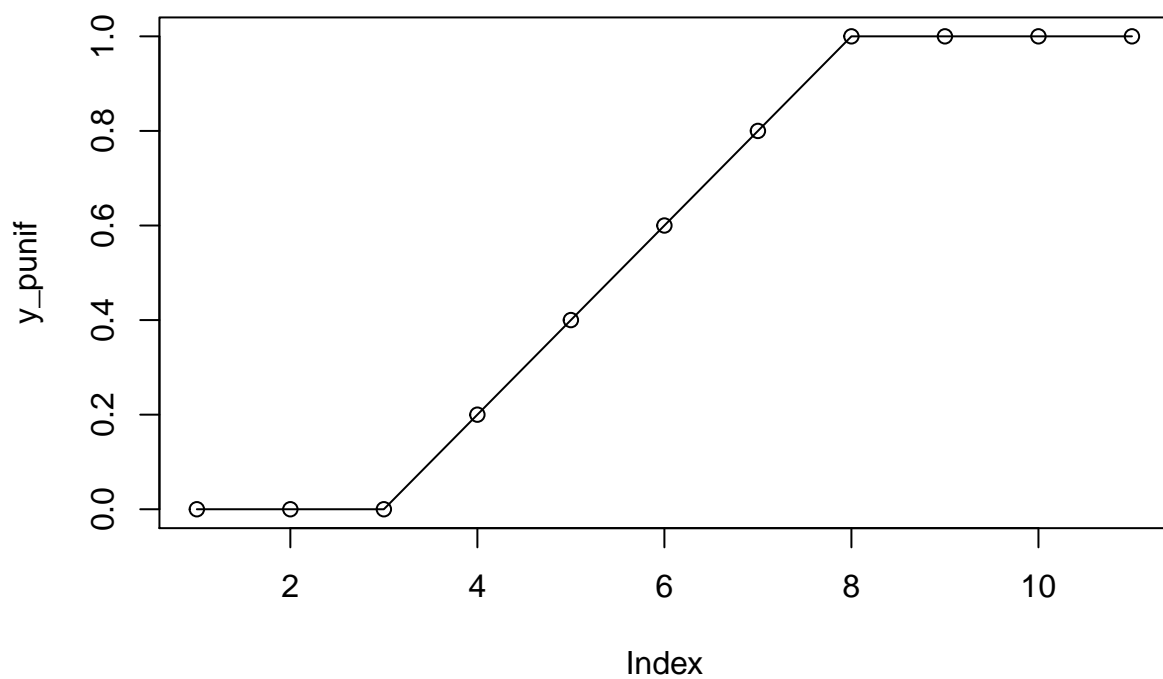
* Quantile Function

```r
x_qunif <- seq(0, 1, by = 0.1)
y_qunif <- qunif(x_qunif, min=2, max=7)
plot(y_qunif, type="o")
```
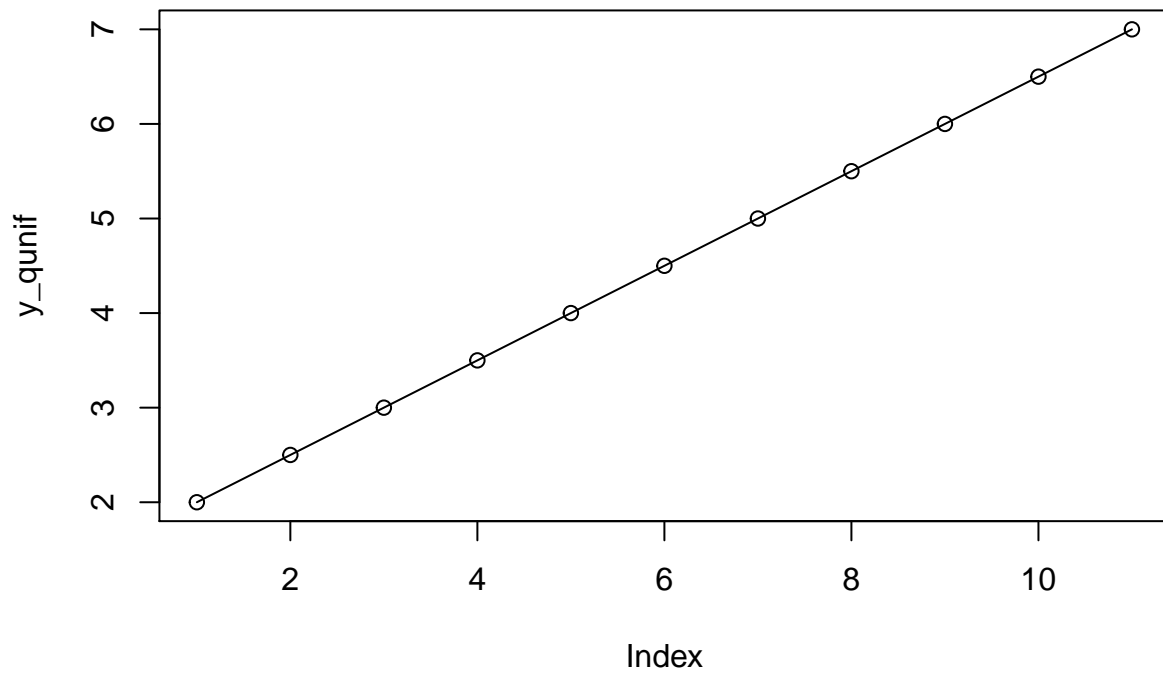
* Generating Random Numbers

```r
set.seed(20211113)
N <- 1000
y_runif <- runif(N, min=2, max=7)
hist(y_runif, breaks = 100, main = "")
```
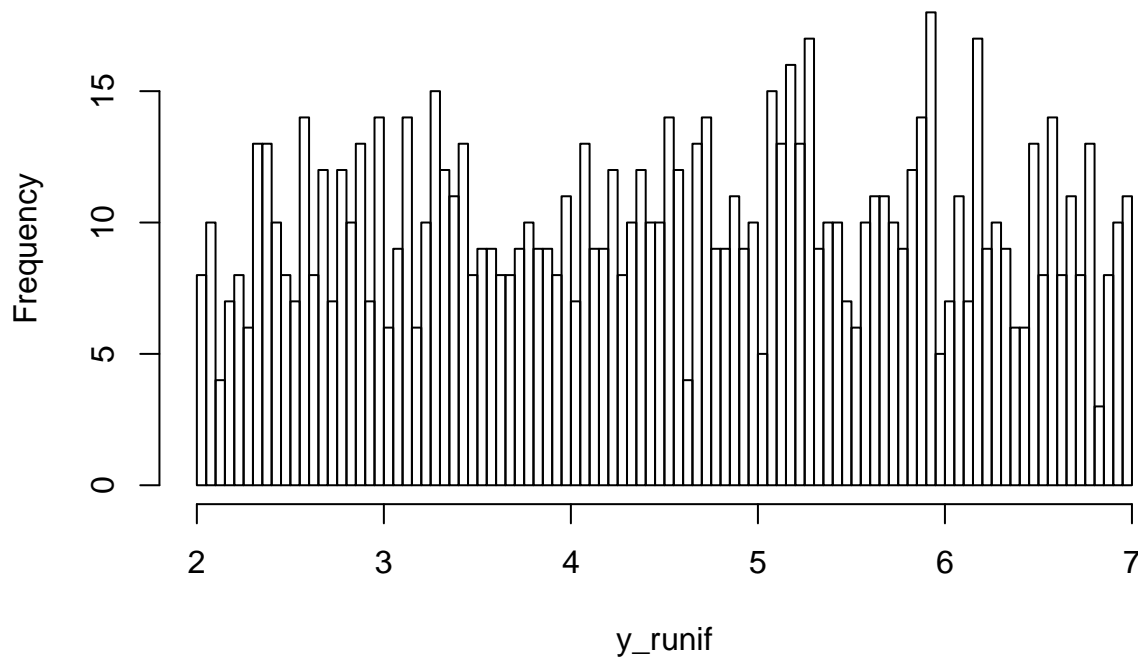
Normal Distribution

- The normal distribution iwth mean μ and variance σ^2 is the continuous distribution with pdf: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$
- The standard normal distribuiton N(0, 1) has zero mean and unit variance, and the standard normal cdf is: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$
- Linear combinations of normal variables are normal, if $X_1, ... X_k$ are independent, X_i $N(\mu_i, \sigma_i^2)$ and $a_1, ... a_k$ are constants, then $Y = a_1 X_1 + ... + a_k X_k$ is normally distributed with mean $\mu = \sum_{i=1}^{k} a_i \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{k} a_i^2 \sigma_i^2$

R code

- pdf

```
x_dnorm <- seq(0, 10, by = 1)
y_dnorm <- dnorm(x_dnorm, mean=2, sd=.5)
plot(y_dnorm, type = "o")
```
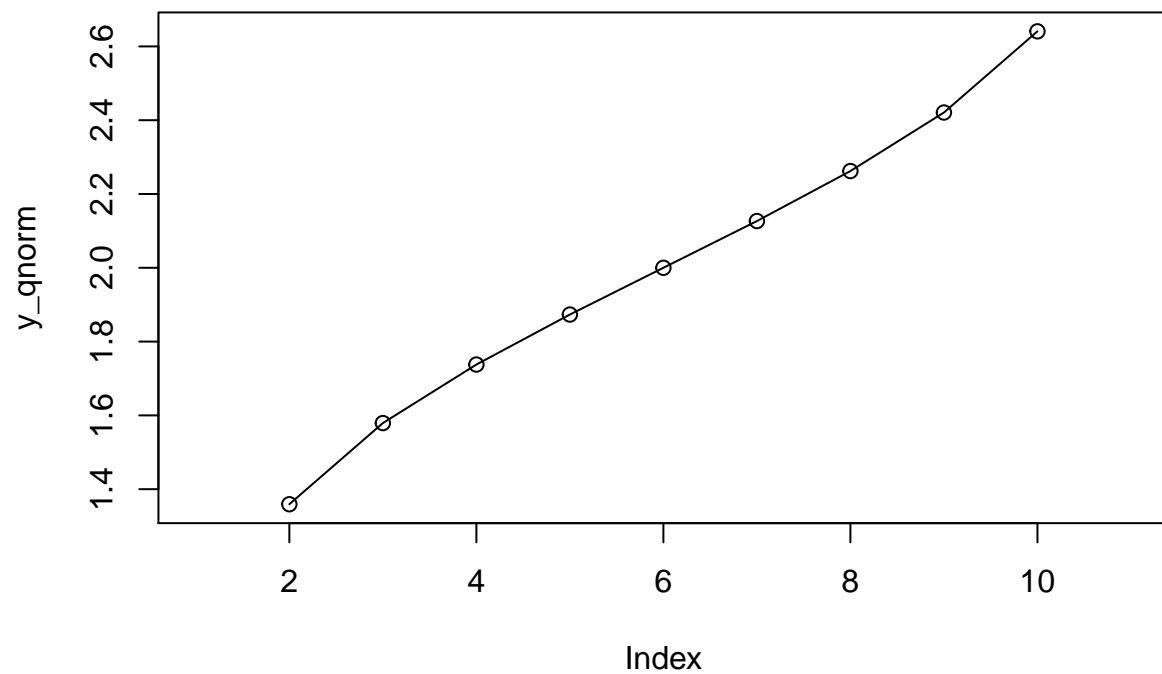
* cdf
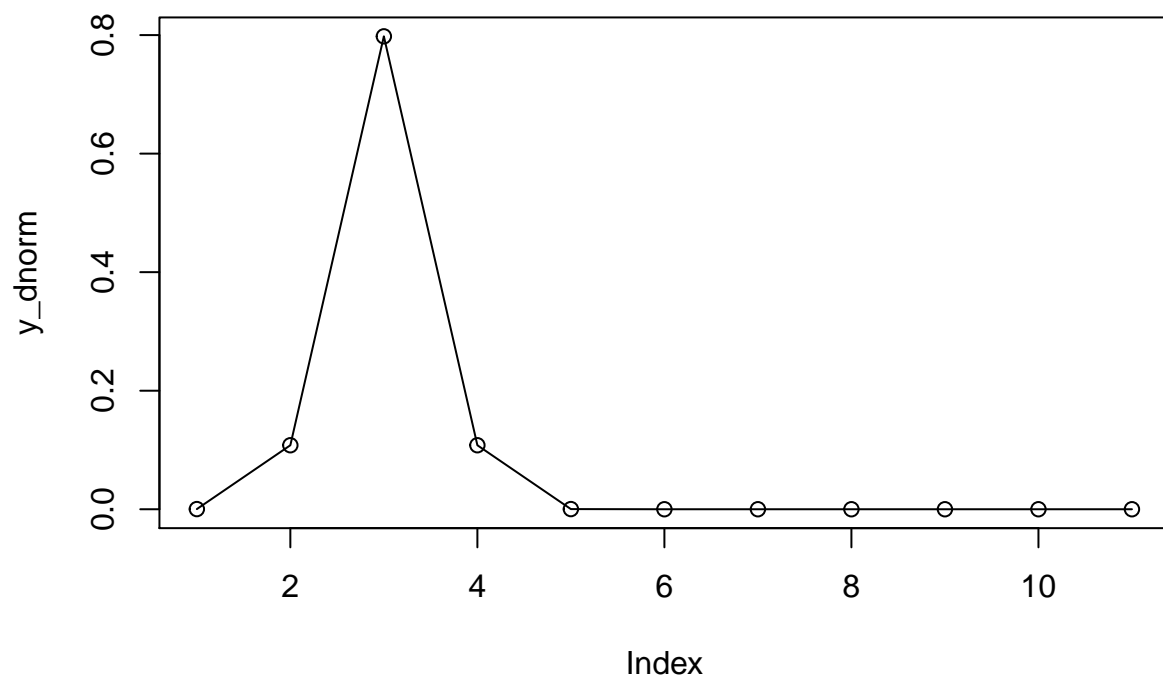
```r
x_pnorm <- seq(0, 10, by=1)
y_pnorm <- pnorm(x_pnorm, mean=2, sd=0.5)
plot(y_pnorm, type="o")
```
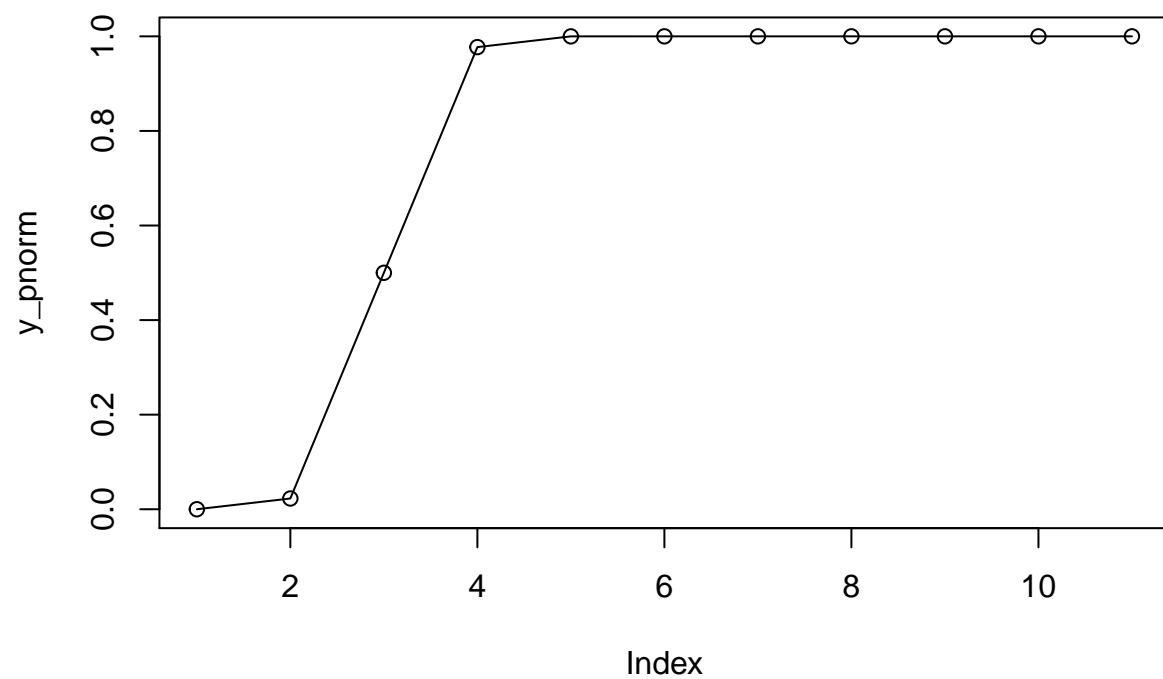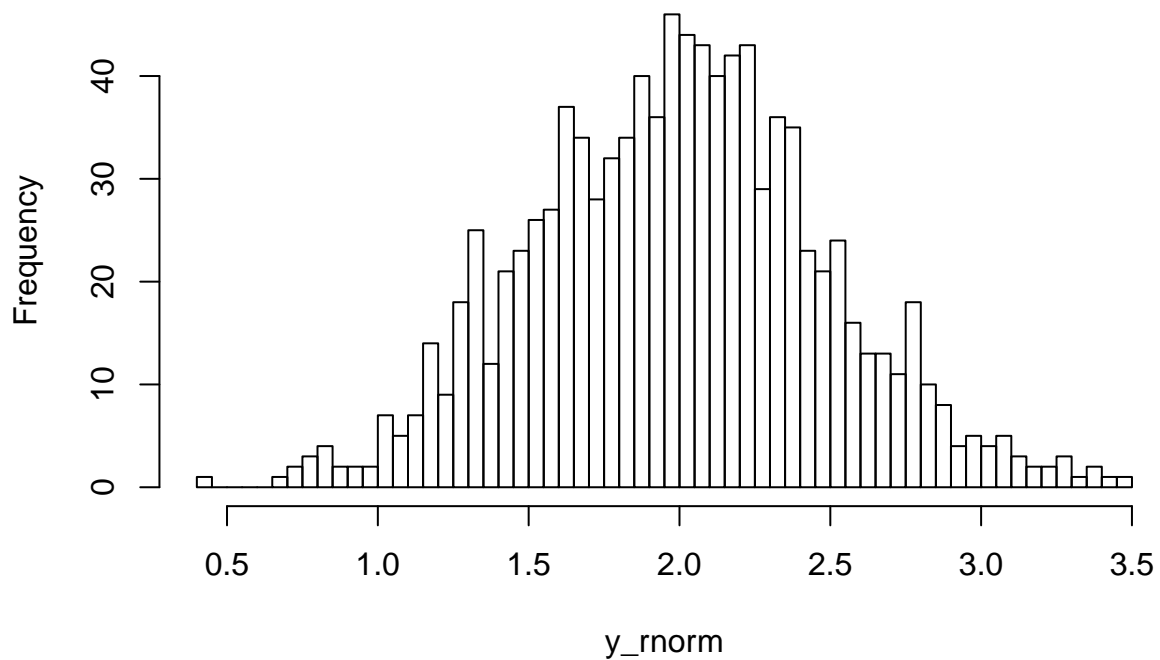
* Quantile Function

```
x_qnorm <- seq(0, 1, by = 0.1)
y_qnorm <- qnorm(x_qnorm, mean=2, sd=0.5)
plot(y_qnorm, type="o")
```

* Generating Random Numbers

```r
set.seed(20211113)
N <- 1000
y_rnorm <- rnorm(N, mean=2, sd=.5)
hist(y_rnorm, breaks = 100, main = "")
```
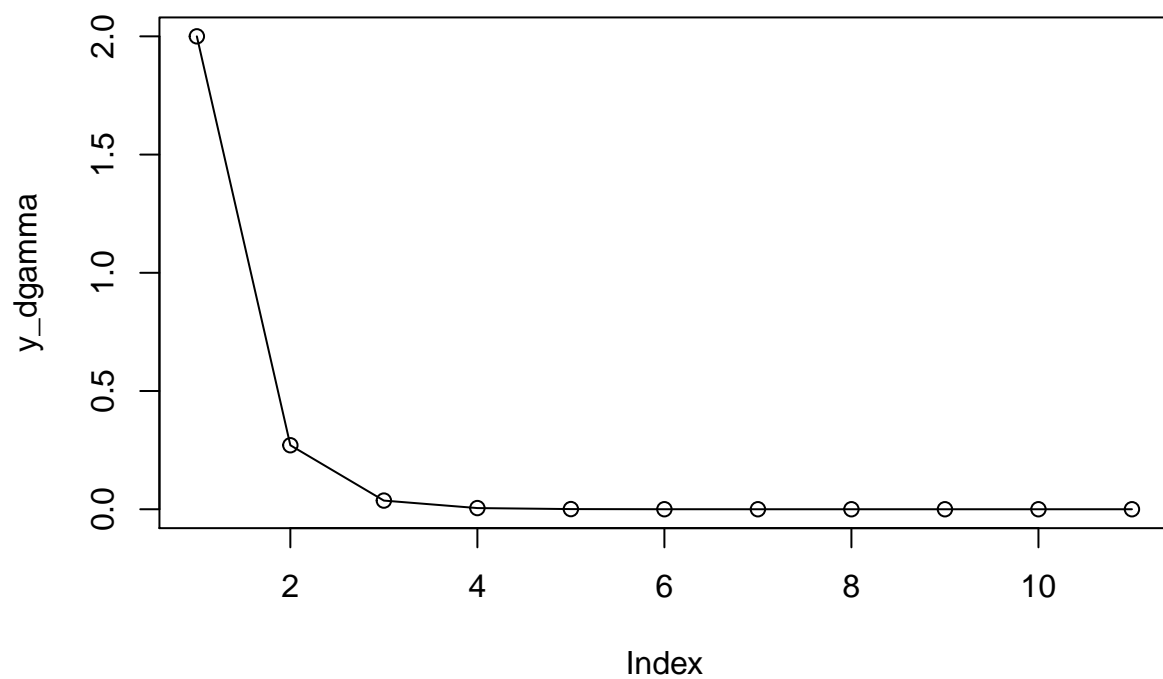
Gamma Distribution

- A random variable X is said to have a gamma distribution with parameter $\alpha > 0$ and $\beta > 0$ with pdf
  $f(x) = \frac{\Gamma(\alpha)\beta^\alpha}{x}^{\alpha-1} e^{-x/\beta} I(0 \le x)$ where $\Gamma(x) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
- It follows that: $E[X] = \alpha\beta$ and $Var(X) = \alpha\beta^2$

R code

- pdf

```
x_dgamma <- seq(0, 10, by = 1)
y_dgamma <- dgamma(x_dgamma, shape=1, rate=2)
plot(y_dgamma, type = "o")
```
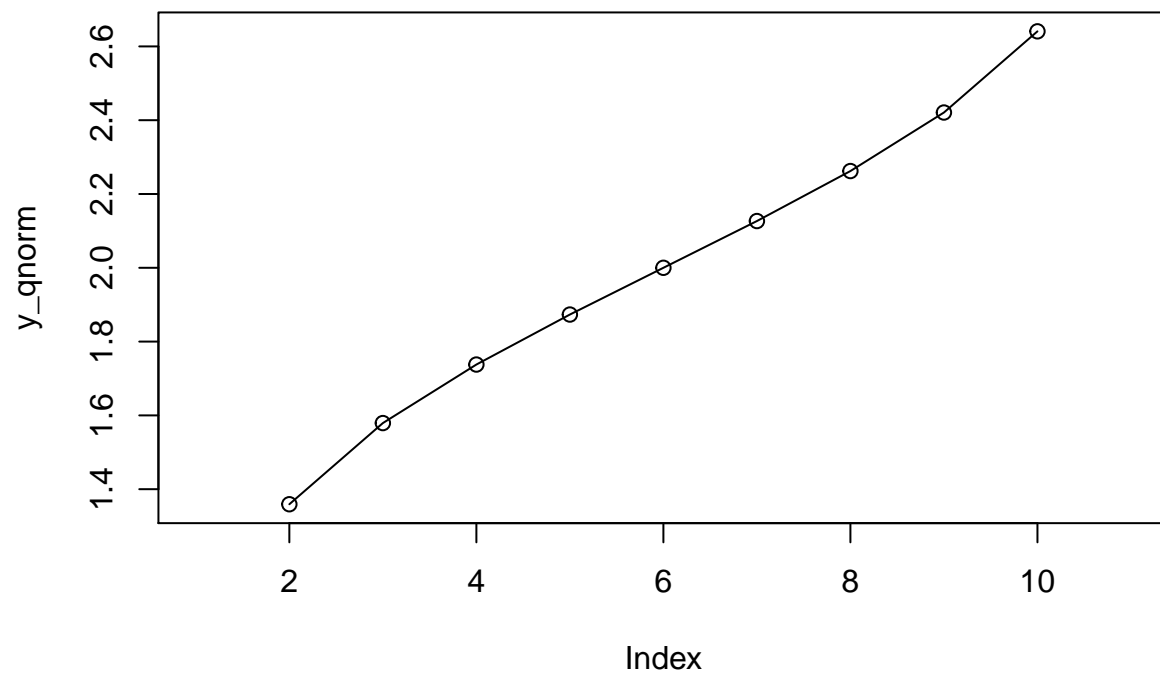
* cdf

```
x_pgamma <- seq(0, 10, by=1)
y_pgamma <- pgamma(x_pgamma, shape=1, rate=2)
plot(y_pgamma, type="o")
```
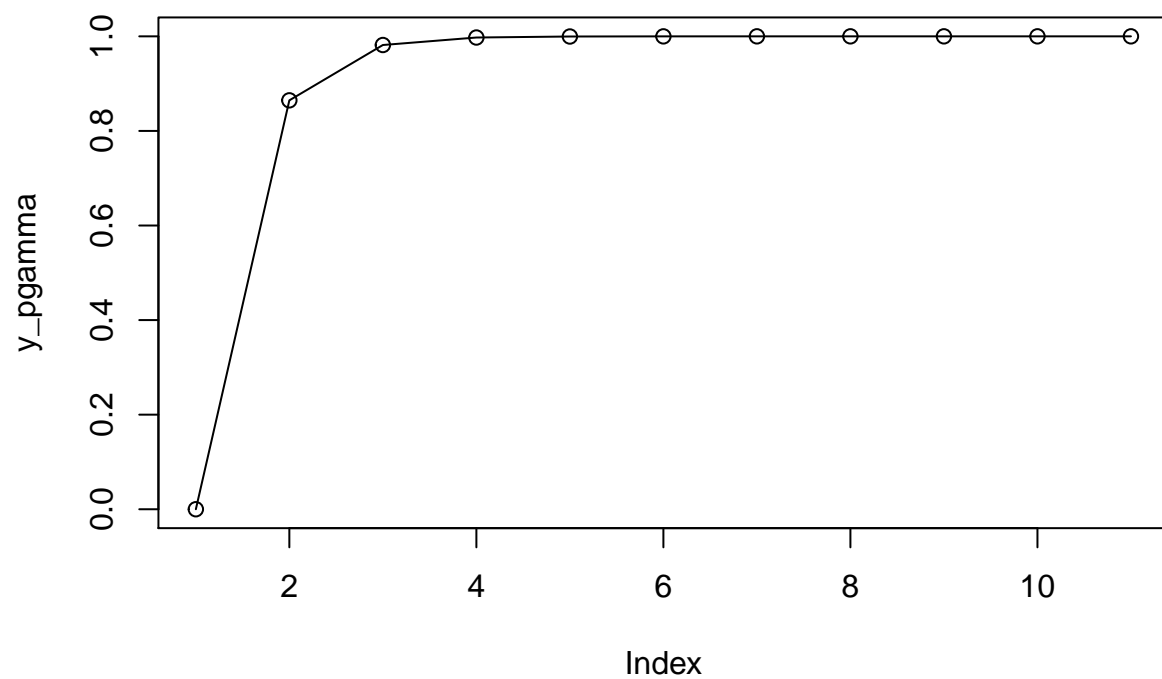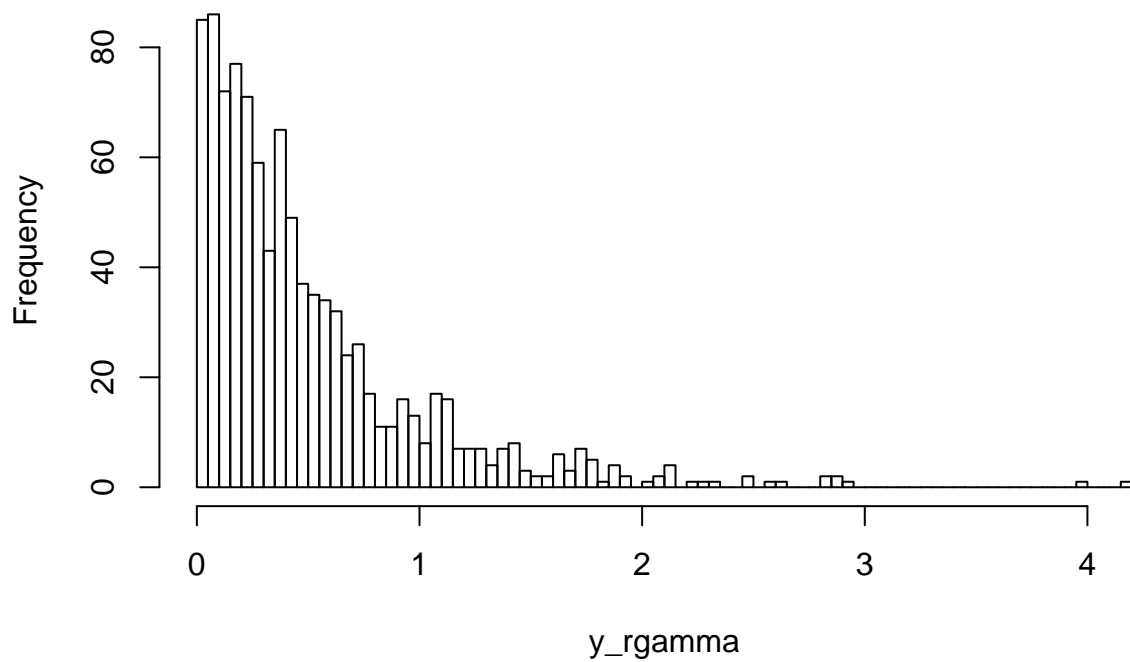
* Quantile Function

```
x_qgamma <- seq(0, 1, by = 0.1)
y_qgamma <- qgamma(x_qgamma, shape=1, rate=2)
plot(y_qnorm, type="o")
```

* Generating Random Numbers

```
set.seed(20211113)
N <- 1000
y_rgamma <- rgamma(N, shape=1, rate=2)
hist(y_rgamma, breaks = 100, main = "")
```

Chi-square Distribution

- A random variable X is said to have a chi-square distribution with v degrees of freedom with pdf $f(x) = \frac{1}{\Gamma(\gamma/2)2^{\gamma/2}}e^{-x/2}$ where γ is a positive integer
- $\chi^2(\gamma)$ is a special case of the gamma distribution, with shape parameter $\gamma/2$
- It follows that $E[X] = \gamma$ and $Var(X) = 2\gamma$
- It $Z_1, ..., Z_\gamma$ are iid standard normal then $Z_1^2 + ... + Z_\gamma^2 \sim \chi^2(\gamma)$
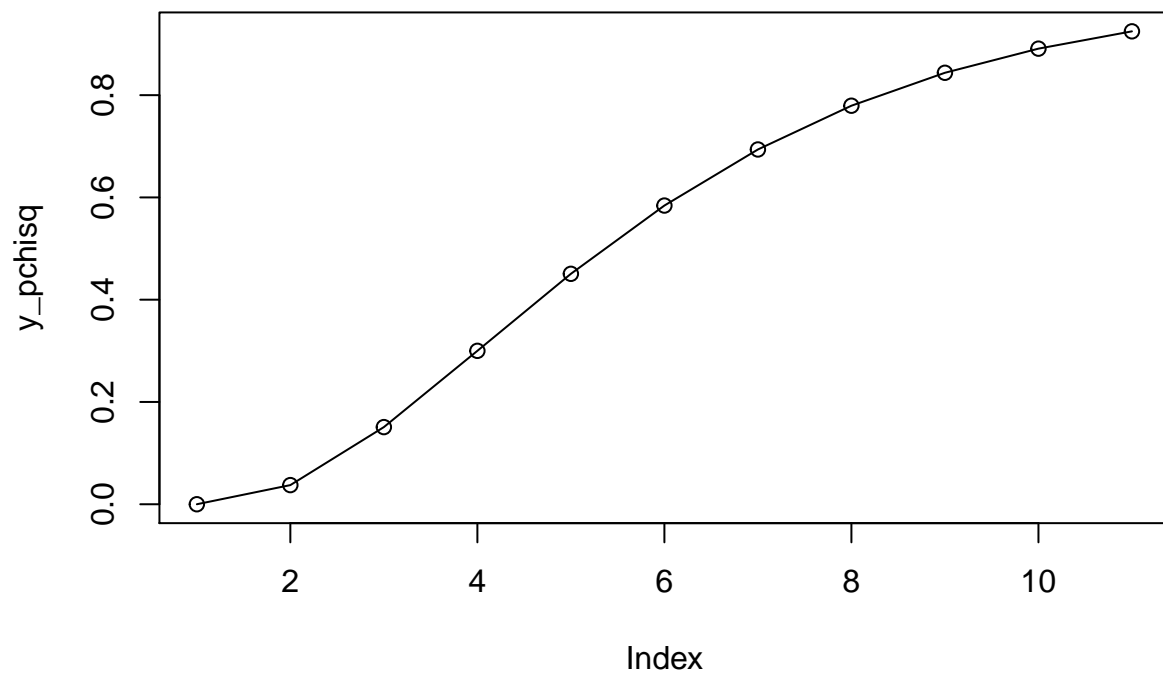
R code

- pdf

```
x_dchisq <- seq(0, 10, by = 1)
y_dchisq <- dchisq(x_dchisq, df=5)
plot(y_dchisq, type = "o")
```
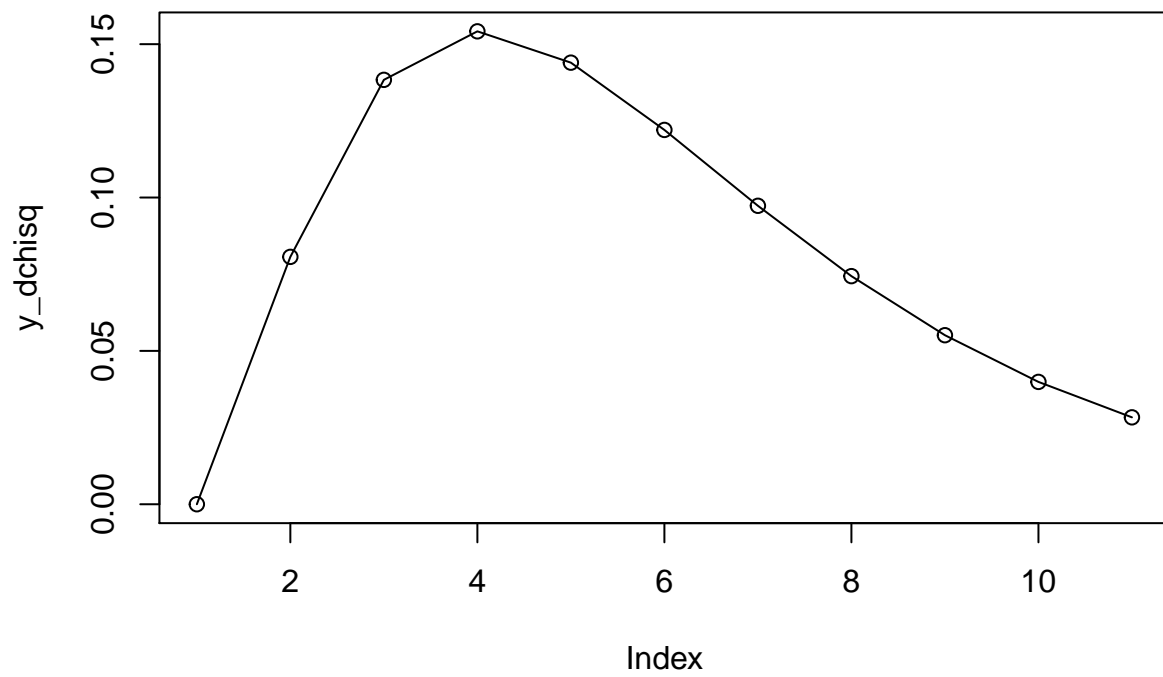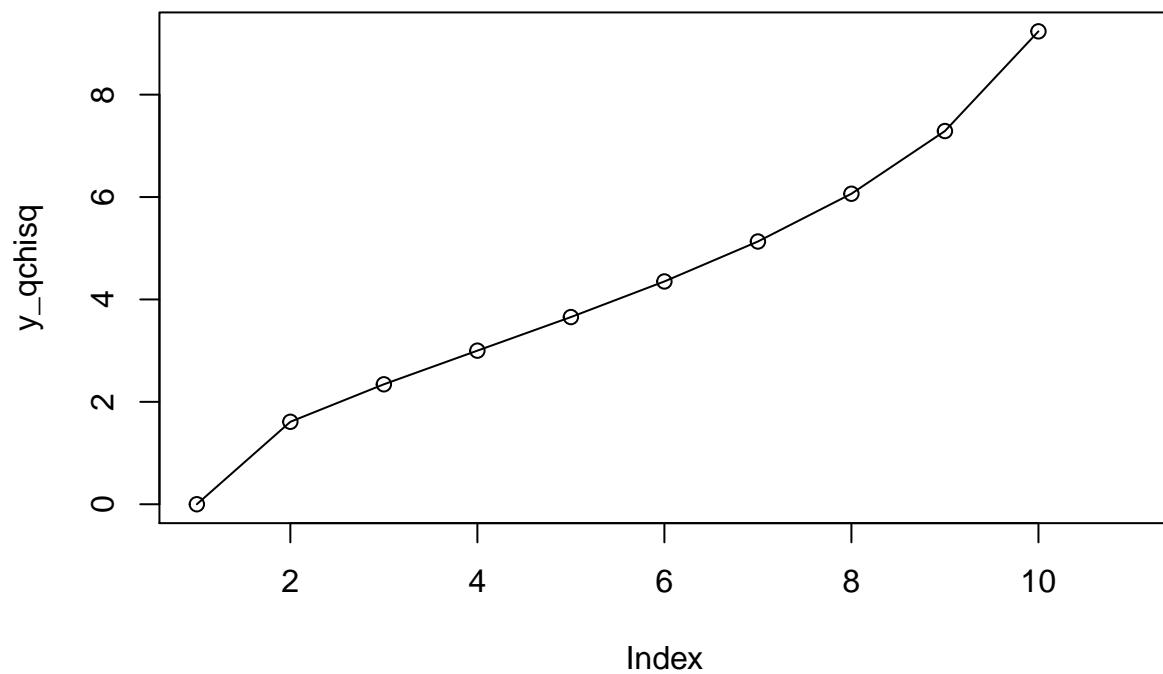
* cdf

```
x_pchisq <- seq(0, 10, by=1)
y_pchisq <- pchisq(x_pgamma, df=5)
plot(y_pchisq, type="o")
```
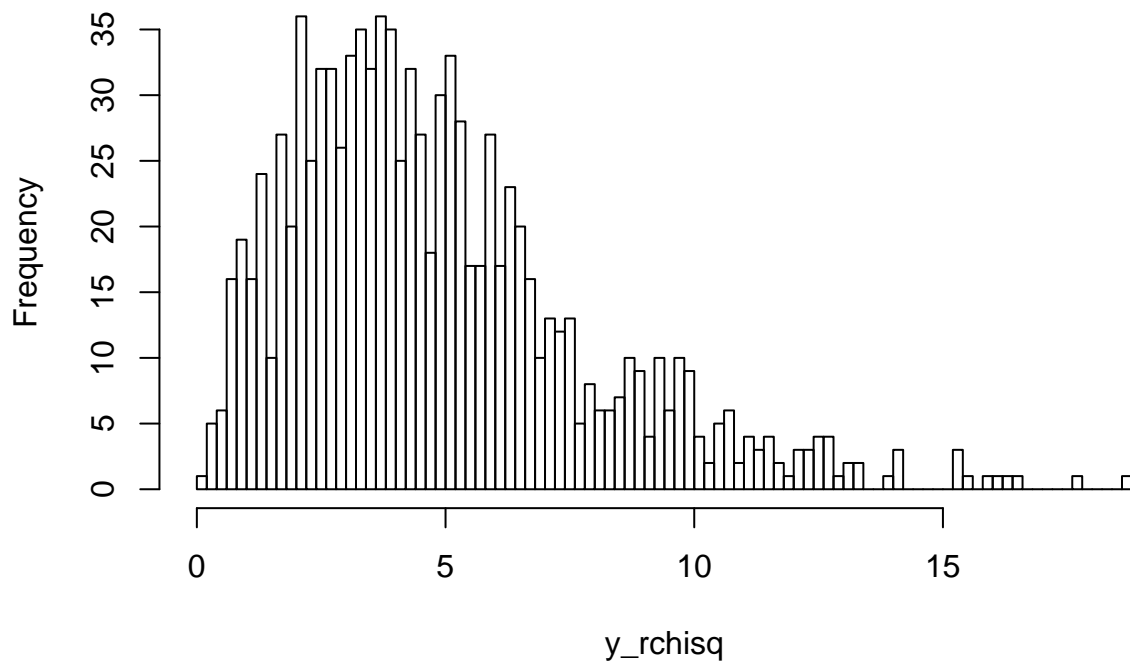
* Quantile Function

```
x_qchisq <- seq(0, 1, by = 0.1)
y_qchisq <- qchisq(x_qchisq, df=5)
plot(y_qchisq, type="o")
```

* Generating Random Numbers

```r
set.seed(20211113)
N <- 1000
y_rchisq <- rchisq(N, df=5)
hist(y_rchisq, breaks = 100, main = "")
```
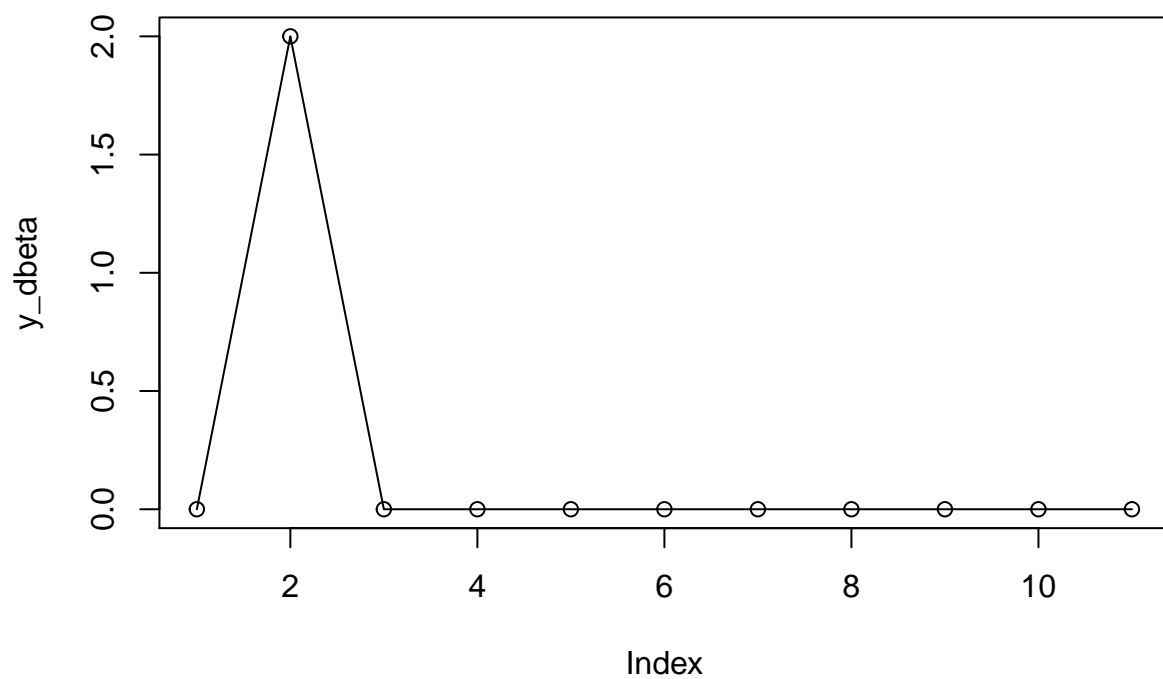
Beta Distribution

- A random variable X is said to have a beta distribution with parameters $\alpha > 0$ and $\beta > 0$ with pdf $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-1)^{\beta-1} I(0 \le x \le 1)$ where $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

- It follows that $E[X] = \frac{\alpha}{\alpha+\beta}$ and $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

- When $\alpha = \beta = 1$, $f(x) = 1, 0 \le x \le 1$. Beta(1, 1) = Unif[0, 1]
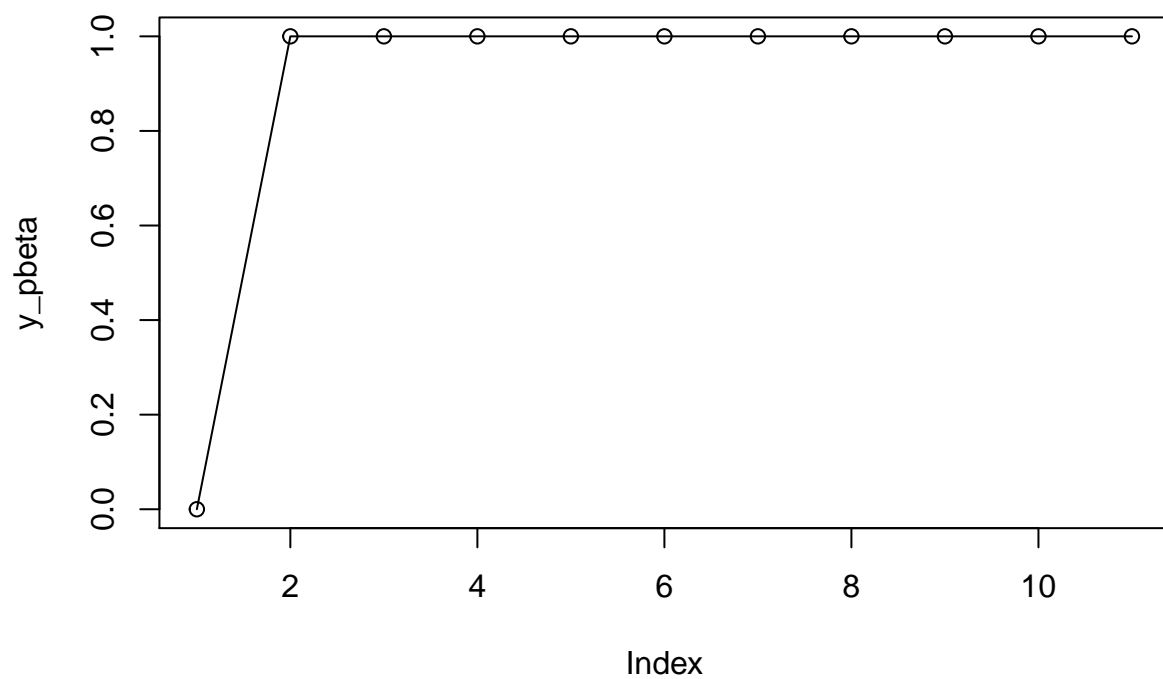
R code

- pdf

```
x_dbeta <- seq(0, 10, by = 1)
y_dbeta <- dbeta(x_dbeta, shape1= 2, shape2= 1)
plot(y_dbeta, type = "o")
```
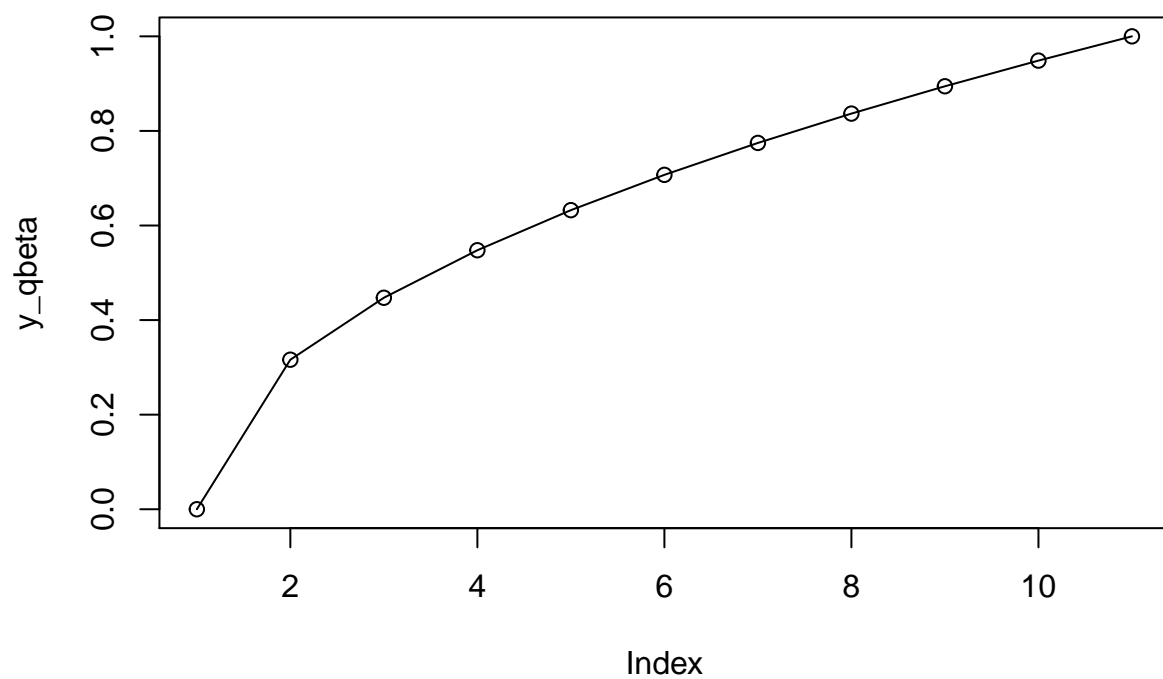
* cdf

```r
x_pbeta <- seq(0, 10, by=1)
y_pbeta <- pbeta(x_pgamma, shape1= 2, shape2= 1)
plot(y_pbeta, type="o")
```
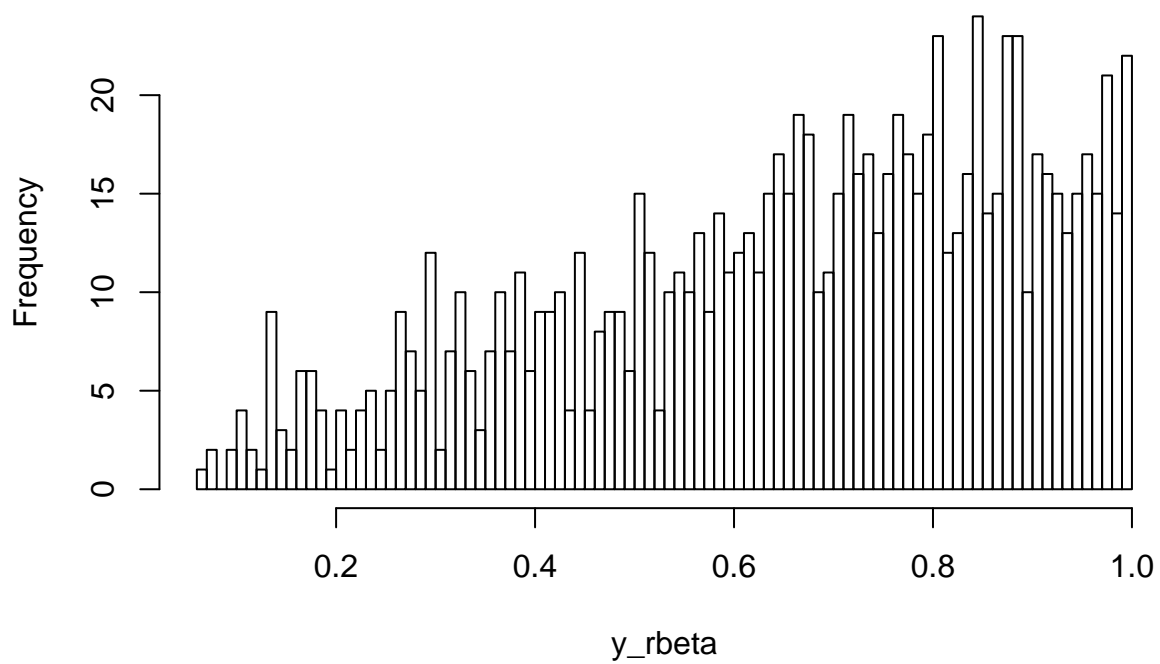
* Quantile Function

```
x_qbeta <- seq(0, 1, by = 0.1)
y_qbeta <- qbeta(x_qbeta, shape1= 2, shape2= 1)
plot(y_qbeta, type="o")
```

* Generating Random Numbers

```r
set.seed(20211113)
N <- 1000
y_rbeta <- rbeta(N, shape1= 2, shape2= 1)
hist(y_rbeta, breaks = 100, main = "")
```

Example

- A A gasoline wholesale distribution has bulk storage tanks that hold fixed suplies and are filled every Monday. Of interest to the whole saler is the proportion of this suppply that is sold during the week. Over many weeks of observation, the distribution found that this proportion could be modeled by a beta distribution with $\alpha = 4$ and $\beta = 2$.

- $P(x \geq 0.9) = \int_{0.9}^{\infty} f(x)dx = \int_{0.9}^{1} 20x^3(1-x)dx = 0.08$