Continuous Random Variables

STAT 211 - 509

2018-10-14

Continuous Random Variables

Let X be a continuous random variable:

- The **probability density function (pdf)** f is defined as

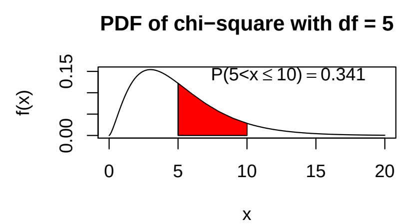
$$P(a < X \leq b) = \int_a^b f(x)dx, \quad a \leq b$$

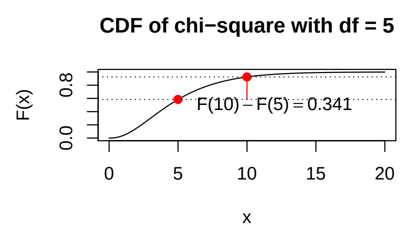
- The **cumulative distribution function (cdf)** F is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx$$

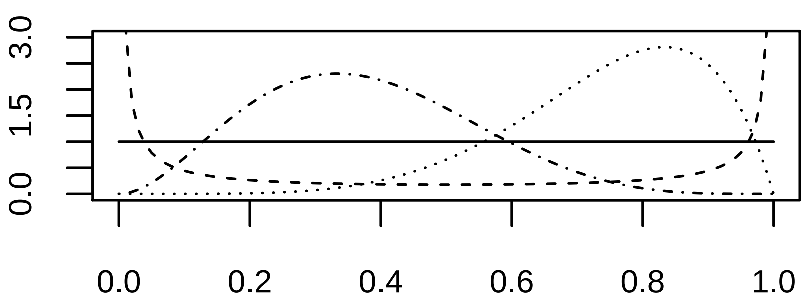
- At every point x at which $f(x)$ is continuous,

$$F'(x) = f(x)$$

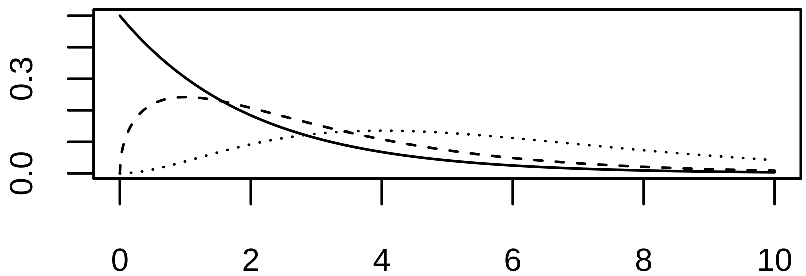




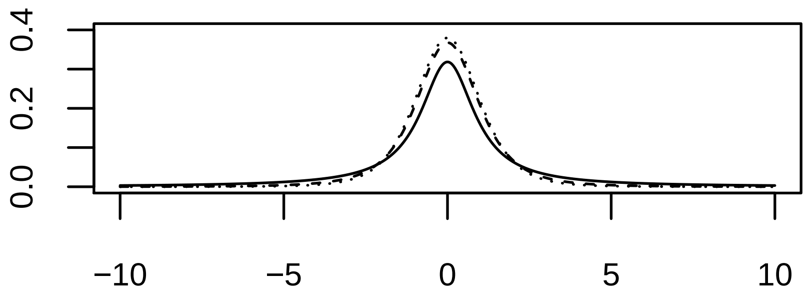
- Beta
  - $f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1]$
  - conjugate prior for binomial distribution parameter p
  - In R: (d | p | q | r)beta



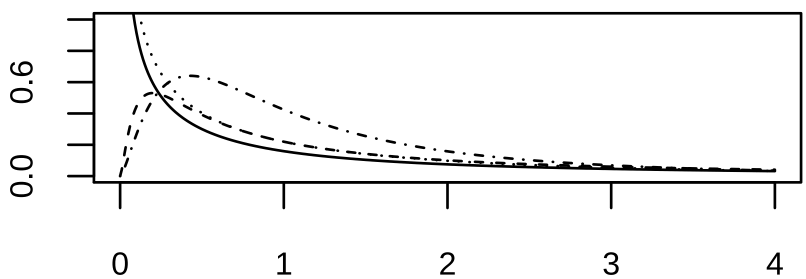
- Chi-square χ^2
  - $f(x;df), \quad x \in [0,\infty)$
  - Chi-square test for goodness of fit
  - In R: (d | p | q | r)chisq



- t
  - $f(x; df), \quad x \in (-\infty, \infty)$
  - t-test for difference in means
  - In R: (d | p | q | r)t



- F
  - $f(x; df1, df2), \quad x \in [0, \infty)$
  - F-test for equality of means in ANOVA
  - In R: (d | p | q | r)f



Continuous distributions in R

Suppose $X \sim \chi^2_2$.
$P(1 \leq X \leq 3)$:

```
pchisq(3, 2) - pchisq(1, 2)
```

```
## [1] 0.3834005
```

$f(3)$:

```
dchisq(3, 2)
```

```
## [1] 0.1115651
```

Sample 10 numbers from X:

```r
rchisq(10, 2)
```

```
##  [1] 0.3102827 3.7648032 3.6090250 1.6723553
##  [5] 2.4450873 2.3167105 1.9800399 0.6147466
##  [9] 0.1892382 0.3144031
```

Expected Values With Continuous RVs

Let X be a continuous random variable:

- The expected value (*mean*) of X is

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

- The expected value of $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- The variance of X is

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

- The moment generating function of is

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Joint and Conditional Distributions

Let X and Y be continuous random variables with marginal pdfs f_X and f_Y and **joint probability density function** $f_{X,Y}$:

- Joint probabilities are double integrals of the joint pdf:

$$P((X,Y) \in A) = \int \int_A f_{X,Y}(x,y) dx dy$$

- The **conditional probability density function** $f_{Y|X=x}(y|x)$ of Y given that $X = x$ is:

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

- X and Y are **independent** if

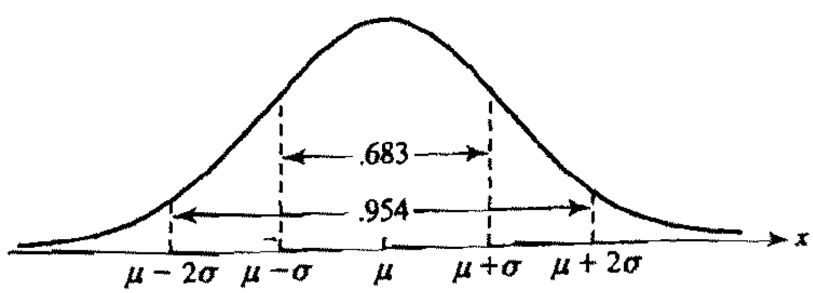
$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

The Normal Distribution

The random variable X has a Normal distribution with parameters μ and σ^2 (written $X \sim N(\mu, \sigma^2)$) if its pdf is

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

- $E[X] = \mu$, $Var(X) = \sigma^2$
- Let $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim N(0,1)$.
- The 68 / 95 / 99.7 rule:
  - $P(-1 < Z < 1) \approx 0.68$
  - $P(-2 < Z < 2) \approx 0.95$
  - $P(-3 < Z < 3) \approx 0.997$

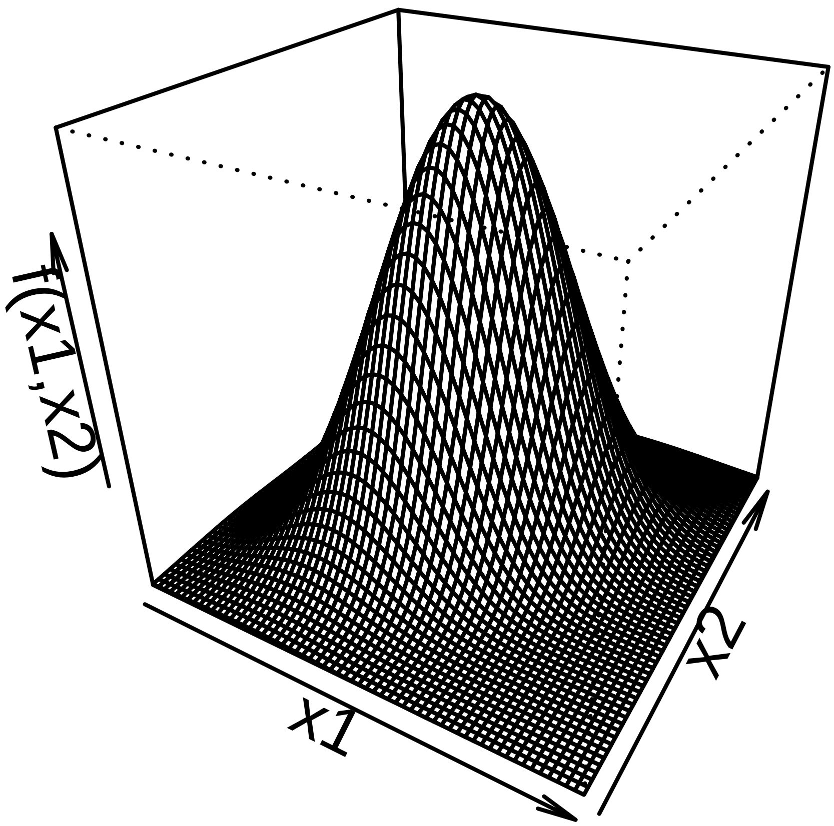


Bivariate normal

If X, Y are bivariate normal, then:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$

where μ_X, μ_Y are the means of X and Y, σ_Y^2, σ_X^2 are the variances, and ρ is the correlation between X and Y.



Sums of Independent Normal RVs

Let $X_1, X_2, \ldots, X_n$ be iid Normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. Let $a_1, a_2, \ldots, a_n$ be arbitrary constants. Then

$$\sum a_i X_i \sim N\Big(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2 \Big)$$

Special case: let $X_1, X_2, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$. Then sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be an IID sample from a distribution with mean μ and variance σ^2. This is not necessarily a Normal distribution. Assume that $\mu < \infty$ and $\sigma^2 < \infty$.

- As $n \to \infty$, the (sampling) distribution of $\bar{X}$ converges to the Normal distribution with mean μ and variance σ^2/n, regardless of the distribution from which the sample was drawn.

- If we have a "large enough" sample size, we can use a Normal distribution to approximate the sampling distribution of $\bar{X}$, regardless of the form of the population distribution.
  - This facilitates mean-based statistical inference.
  - Common rule of thumb for "large enough" is 30, although the required sample size varies depending on the situation.