Statistics

Sampled from

Morris H. DeGroot & Mark J. Schervish, “Probability and Statistics”, 3rd Edition, Addison Wesley

Probability: Continuous distribution and variables

- Continuous distributions
  - Random variables
  - Probability density function
  - Uniform, normal and exponential distributions
  - Expectations and variance
  - Law of large numbers
  - Central limit theorem
  - Probability density functions of more than one variable
  - Rejection and transformation methods for sampling distributions (section)

Statistics

- Estimators: mean, standard deviation
- Maximum likelihood
- Confidence intervals
- χ^2 statistics
- Regression
- Goodness of fit

Continuous random variables

- A random variable *X* is a real value function defined on a sample space *S.*
- *X* is a continuous random variable if a non-negative function *f,* defined on the real line, exists such that an integral over the domain *A* is the probability that *X* takes a value in domain *A.* (*A* is, for example, the interval [a,b])

$$\Pr\left(a < X < b\right) = \int_a^b f\left(x\right) dx$$

Probability density function

- *f* is called probability density function (p.d.f.). Note that the unit of the pdf below are of 1/length, only after the multiplication with a length element we get probability
- For every p.d.f. we have

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Examples of p.d.fs

- A car is driving in a circle at a constant speed. What is the probability that it will be found in the interval between 1 and 2 radians?
- A computer is generating with equal probability density, random numbers between 0 and 1. What is the probability of obtaining 0.75?
- Protein folds at a constant rate (the probability that a protein will fold at the time interval $[t,t+dt]$ is a constant αdt). If we have at time zero N_0 protein molecules, what is the probability that all protein molecules will fold after time t' ?

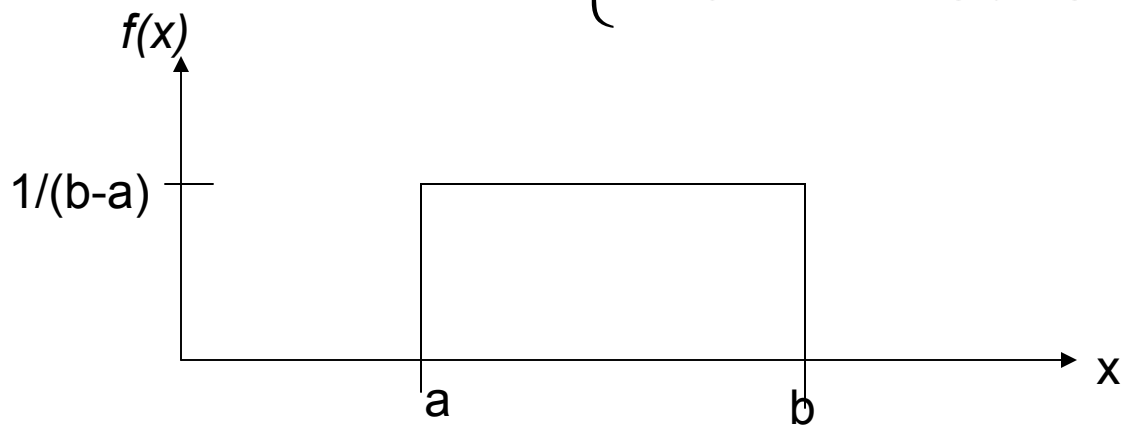
Uniform distribution on an interval

- Consider an experiment in which a point X is selected from an interval $S = \{x : a \le x \le b\}$ in such a way that the probability of finding X at a given interval is proportional to the interval length (hence the p.d.f. is a constant). This distribution is called the uniform distribution We must have for this distribution

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{a}^{b} f(x)\,dx = 1$$

Uniform distribution (continue)

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



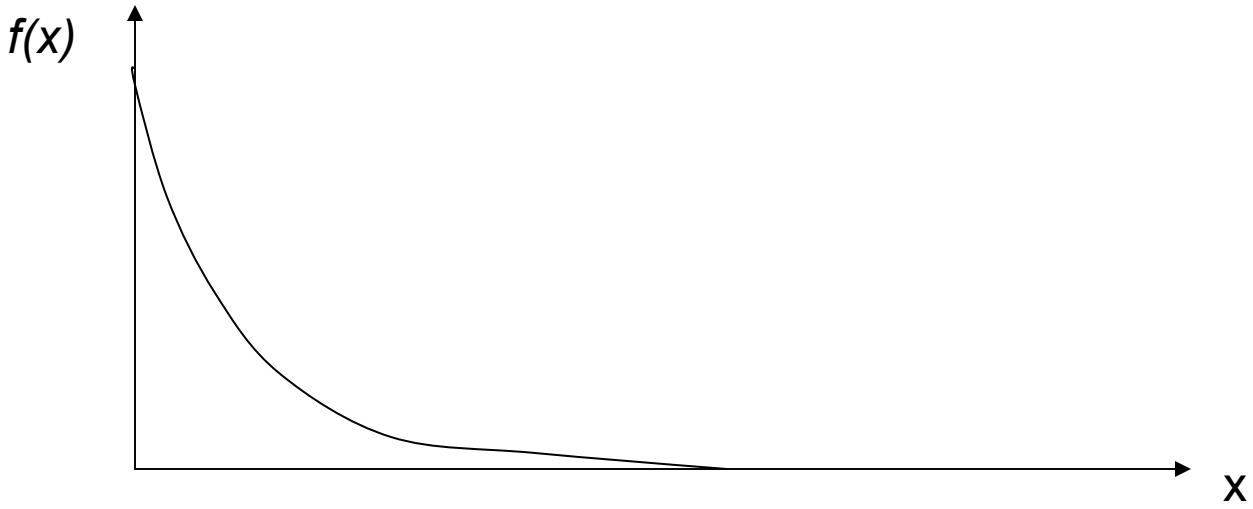
Examples of simple distributions

- *X* is a random variable distributed uniformly on a circle of radius *a*. Find *f(x)*
- Check that the following function satisfies the conditions to be a p.d.f.

$$f(x) = \begin{cases} \frac{2}{3} x^{-1/3} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

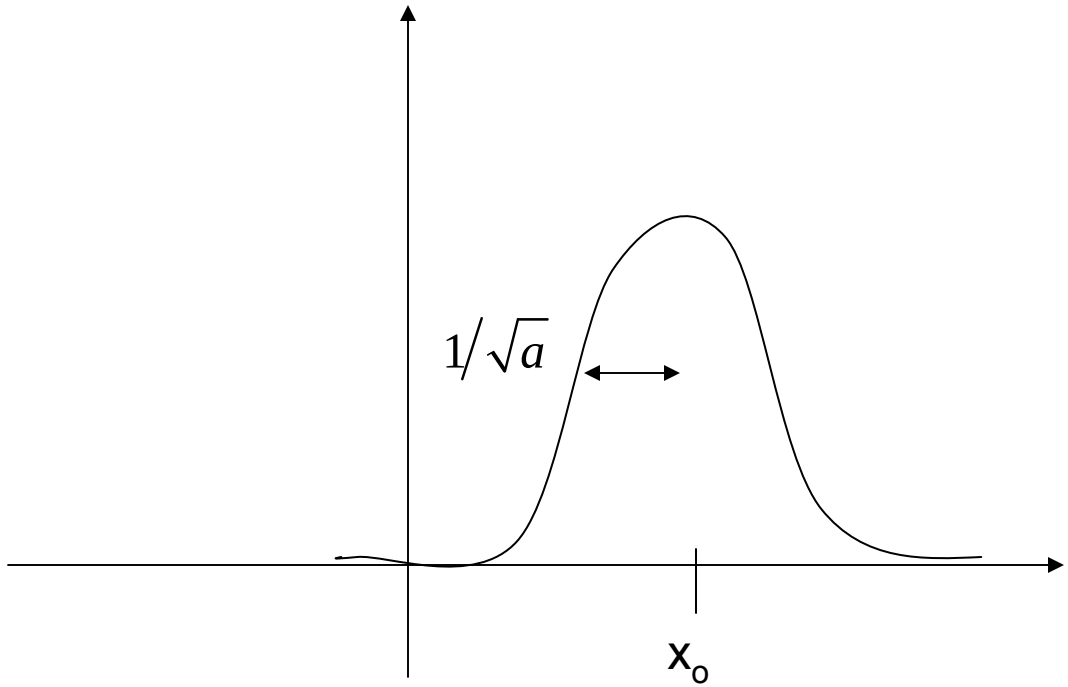
Exponential distribution

$$f(x) = \left[a \exp(-ax) \quad 0 < x < \infty \right]$$



Normal distribution

$$f(x) = \left[\left(\frac{a}{\pi} \right)^{1/2} \exp\left(-a\left(x - x_0 \right)^2 \right) \quad -\infty < x < \infty \right]$$



Continuous distribution functions

$$\text{defined as } F(x) = \Pr(X \le x) \qquad \text{for } -\infty < x < \infty$$

F(x) is a monotonic non decreasing function of *x (can you show it?),* that can be written in terms of its corresponding p.d.f.

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

or

$$\frac{dF}{dx} = f(x)$$

Distribution function: Example

$$f(x) = \begin{bmatrix} a\exp(-ax) & \text{for } 0 < x < \infty \\ 0 & \text{otherwise} \end{bmatrix}$$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{x} a\exp(-ax)\,dx = \left[-\exp(-ax)\right]_0^x = 1 - \exp(-ax)$$

Expectation

- For a random variable *X* with a p.d.f. *f(x)* the expectation *E(X)* is defined

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

The expectation exists if and only if the integral is absolutely converged, i.e.

$$\int_{-\infty}^{\infty} |x|\, f(x) dx < \infty$$

Expectation (example)

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot 2x \cdot dx = 2\left[\frac{x^3}{3}\right]_0^1 = \frac{2}{3}$$

Even if the p.d.f, satisfies the requirements, it is not obvious that the expectation exists (next slide)

The Cauchy p.d.f.

$$f(x) = \left[\frac{1}{\pi\left(1+x^2\right)} \quad -\infty < x < \infty \right] \qquad \left(f(x) \geq 0 \right)$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\pi\left(1+x^2\right)} dx = \frac{1}{\pi} \arctan(x) \Big|_{-\infty}^{x} = \frac{1}{\pi}\left(\arctan(x) - \left(-\frac{\pi}{2} \right) \right)$$

$$F(\infty) = \frac{1}{\pi}\left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1 \qquad \left(\int_{-\infty}^{\infty} \frac{1}{\pi\left(1+x^2\right)} dx = 1 \right)$$

Cauchy distribution: Expectation

Test for existence of expectation

$$E(X) = \int_{-\infty}^{\infty} |x| \cdot f(x) \cdot dx = \int_{-\infty}^{\infty} |x| \frac{1}{\pi\left(1+x^2\right)} dx \to \infty$$

Expectation does not exist for the Cauchy distribution.

Some properties of expectations

- Expectation is linear

$$E\left(aX + bY\right) = aE\left(X\right) + bE\left(Y\right)$$

- If the random variables *X* and *Y* are independent $\left(f\left(x,y\right) = f\left(x\right)f\left(y\right)\right)$ then

$$E\left(X \cdot Y\right) = E\left(X\right) \cdot E\left(Y\right)$$

Expectation of a function

- Is essentially the same as the expectation of a variable

$$E\left(r\left(x\right)\right)=\int_{-\infty}^{\infty} r\cdot g\left(r\right)dr=\int_{-\infty}^{\infty} r\left(x\right)\cdot f\left(x\right)\cdot dx$$

of special interest is the expectation value of moments

$$\text{variance}\equiv E\left(X^{2}\right)-\left[E\left(X\right)\right]^{2}=\int_{-\infty}^{\infty} x^{2}\cdot f\left(x\right)\cdot dx-\left[\int_{-\infty}^{\infty} x\cdot f\left(x\right)\cdot dx\right]^{2}$$

Can you show that the variance is always non-negative?

Functions of several random variables

- We consider a p.d.f.

$$f\left(x_1,...,x_n\right)$$

of several random variables

$$X_1,...,X_n$$

The p.d.f. satisfies (of course)

$$f\left(x_1,...,x_n\right)\geq 0$$

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}...\int_{a_n}^{b_n} f\left(x_1,...,x_n\right)dx_1...dx_n = 1$$

Expectation of function of several variables

- Similarly to one variable case, expectations of functions with several variables are computes

$$E\left(Y = r\left(x_1,...,x_n\right)\right) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} r\left(x_1,...,x_n\right) \cdot f\left(x_1,...,x_n\right) dx_1 ... dx_n$$

Example: expectation of more than one variable

$$f(x,y)=\begin{cases}1 & \text{for } (x,y)\in S\\ 0 & \text{otherwise}\end{cases}$$

S is a square: $0<x<1\ 0<y<1$

$$E\left(X^2+Y^2\right)=\int_0^1\int_0^1\left(x^2+y^2\right)f(x,y)\cdot dx\cdot dy$$

$$=\int_0^1\int_0^1\left(x^2+y^2\right)dx\cdot dy=\frac{2}{3}$$

Markov Inequality

- *X* is a random variable such that

$$\Pr\left(X \geq 0\right) = 1$$

- For every *t>0*

$$\Pr\left(X \geq t\right) \leq \frac{E\left(X\right)}{t}$$

- Prove it
- Why *E(X)>t* is not interesting?

Chebyshev Inequality
is a special case of the Markov inequality

- *X* is a random variable for which the variance exists. For *t>0*

$$\Pr\left(\left[X - E(X)\right]^2 \geq t^2\right) \leq \frac{\mathrm{var}(X)}{t^2}$$

- Substitute

$$Y = \left[X - E(X)\right]^2 \quad \rightarrow E(Y) = \mathrm{var}(X) \quad \text{and } t^2 \text{ by } t$$

to obtain the Markov inequality

The law of large numbers I

- Consider a set of *N* random variables $X_1,...X_n$

i.i.d. Each of the random variables has mean (expectation value) μ and variance σ^2

- The arithmetic average of *n* samples is defined $\bar{X}_n = \frac{1}{n}\left(X_1 + ... + X_n\right)$. It defines a new random variable that we call the sample mean
- The expectation value of the sample mean

$$E\left(\bar{X}_n\right) = \frac{1}{n}\sum_i E\left(X_i\right) = \frac{1}{n}\cdot n\mu = \mu$$

The Law of Large Numbers II

- The variance of $\bar{X}_n$

$$\mathrm{var}\left(\bar{X}_n\right) = E\left(\bar{X}_n^2 - E^2\left(\bar{X}_n\right)\right) = \frac{1}{n^2}\sum_{i,j} E\left(X_i X_j\right) - \frac{1}{n^2}\left[\sum_i E\left(X_i\right)\right]^2$$

Since X_i and X_j are independent for $i \neq j$ $E\left(X_i X_j\right) = E\left(X_i\right)E\left(X_j\right)$

$$\mathrm{var}\left(\bar{X}_n\right) = \frac{1}{n^2}\left[n \cdot E\left(X^2\right) + \left(n^2 - n\right) \cdot E^2\left(X\right) - n^2 E^2\left(X\right)\right]$$

$$\mathrm{var}\left(\bar{X}_n\right) = \left(E\left(X^2\right) - E^2\left(X\right)\right)\Big/ n = \mathrm{var}\left(X\right)\Big/ n$$

Which means that the variance is decreasing linearly with the number of sampled points

Law of Large numbers III

Chebyshev Inequality:

$$1-\Pr\left(\left(\bar{X}_n-\mu\right)^2\geq\varepsilon^2\right)=\Pr\left(\left(\bar{X}_n-\mu\right)^2<\varepsilon^2\right)\geq 1-\frac{\operatorname{var}\left(X\right)}{n\varepsilon^2}\quad\text{for }\varepsilon\geq 0$$

$$\Rightarrow \bar{X}_n\to\mu$$

Central Limit Theorem

- Statement without proof:
- Given a set of random variables $X_1, ..., X_n$ with mean μ_i and variance σ^2_i we define a new random variable

$$Y_n = \frac{\sum_{i=1,...,n} X_i}{\left(\sum_{i=1,...,n} \sigma_i^2 \right)^{1/2}}$$

- For very large *n*, the distribution of $\sum_{i=1,...,n} X_i$ is normal with mean $\sum_{i=1,..,n} \mu_i$ and variance $\sum_{i=1,...,n} \sigma_i^2$

Statistical Inference

- Data generated from unknown probability distribution and statement on the unknown distribution are warranted. Determine parameters (e.g. β for exponential distribution, μ and σ for normal distribution)
- Prediction of new experiments

Estimation of parameters

- Notation: *f(x|θ)* is the probability density of sampling *x* given (conditioned on) parameters *θ.*
- For a set of *n* independent and identically distributed samples the probability density is:

$$f\left(x_1,...,x_n \mid \theta\right) = \prod_{i=1,...,n} f\left(x_i \mid \theta\right) \equiv f\left(\mathbf{x} \mid \theta\right)$$

- However, what we want to determine now are the parameters… For example assuming the distribution is normal, we seek the mean μ and the variance σ^2

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

Bayesian arguments

- What we want is the function $f\left(\theta \mid \mathbf{x}\right)$ given a set of observations **x**, what is the probability that the set of parameters is θ?
- Bayesian statistics: Think of the parameters like other random variables with probability $\xi(\theta)$.

The joint probability $f\left(\mathbf{x},\theta\right) \equiv f\left(\mathbf{x} \mid \theta\right)\xi\left(\theta\right)$ is also

$$f\left(\mathbf{x},\theta\right) \equiv f\left(\theta \mid \mathbf{x}\right)g\left(\mathbf{x}\right)$$

The likelihood function

- We can formally write $\xi(\theta|\mathbf{x}) = \dfrac{f(\mathbf{x}|\theta)\xi(\theta)}{g(\mathbf{x})}$

which is the probability of having a particular set of parameter for the p.d.f provided a set of observation (what we wanted). Note that our prime interest here is in the parameter set θ and the samples of x is given. Since $g(x)$ is independent of θ we can write the likelihood function

$$\xi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\xi(\theta)$$

Example: Likelihood function I

- Consider the exponential distribution

$$f\left(x \mid \beta\right) = \left\{ \begin{array}{cc} \beta \exp\left[-\beta x\right] & \text{for } x > 0 \\ 0 & \text{otherwise} \end{array} \right\}$$

$$f\left(\mathbf{x} \mid \beta\right) = \left\{ \begin{array}{cc} \beta^n \exp\left[-\beta \sum_{i=1,\ldots,n} x_i \right] & \text{for } x > 0 \\ 0 & \text{otherwise} \end{array} \right\}$$

- And assume the p.d.f. of the parameter β is a Gaussian with a mean and variance of 1.

$$\xi\left(\beta\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\beta^2}{2} \right)$$

Example: Likelihood function II

$$\xi(\beta \mid \mathbf{x}) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{\beta^2}{2}\right) \beta^n \exp\left[-\beta \sum_{i=1,\ldots,n} x_i\right]$$

Maximum Likelihood

We look for a maximum of the function $L(\theta) = \log\left(f_n(\mathbf{x} \mid \theta)\right)$ as a function of the parameters θ

As a concrete example we consider the normal distribution

$$L(\theta) = \log\left[f_n(\mathbf{x} \mid \mu, \theta)\right]$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1,\ldots,n}(x_i - \mu)^2$$

To find the most likely set of parameters we determine the maximum of *L(θ)*

Maximum of $L(\theta)$ for normal distribution

$$\frac{dL}{d\mu} = 0 = -\frac{1}{2\sigma^2}\sum_{i=1,\ldots,n} 2\left(x_i - \mu\right) = -\frac{1}{\sigma^2}\left(\sum_{i=1,\ldots,n} x_i - n\mu\right)$$

$$\Rightarrow \mu = \frac{1}{n}\sum_{i=1,\ldots,n} x_i$$

$$\frac{dL}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1,\ldots,n}\left(x_i - \mu\right)^2 = 0$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1,\ldots,n}\left(x_i - \mu\right)^2$$

Determine a most likely parameter for the uniform distribution

$$f\left(x\,|\,\theta\right)=\begin{cases}\dfrac{1}{\theta} & \text{for } 0\le x\le\theta\\ 0 & \text{otherwise}\end{cases}$$

$$f\left(\mathbf{x}\,|\,\theta\right)=\begin{cases}\dfrac{1}{\theta^{n}} & \text{for } 0\le x_i\le\theta\ \left(i=1,\ldots,n\right)\\ 0 & \text{otherwise}\end{cases}$$

It is clear that θ must be larger than all the x_i and at the same time maximizes the monotonically decreasing function $1/\theta^n$, hence

$$\theta=\max\left[x_1,\ldots,x_n\right]$$

Potential problems in maximum likelihood procedure

• Value of θ is underestimated (note that θ should be larger than all *x*, not only the ones we sample so far)

•No guarantee that a solution exists for the distribution below θ must be large than any *x* but at the same time equal to the maximal *x*. This is not possible and hence, no solution

$$f\left(x\,|\,\theta\right)=\begin{Bmatrix}\frac{1}{\theta} & \text{for } 0<x<\theta\\ 0 & \text{otherwise}\end{Bmatrix}$$

•The solution is not necessarily unique

$$f_n\left(\mathbf{x}\,|\,\theta\right)=\begin{Bmatrix}1 & \text{for } \theta\le x_i\le\theta+1\ \left(\text{i=1,...,n}\right)\\ 0 & \text{otherwise}\end{Bmatrix}$$

$$\Rightarrow f_n\left(\mathbf{x}\,|\,\theta\right)=\begin{Bmatrix}1 & \text{for } \max\left(x_1,...,x_n\right)-1\le\theta\le\min\left(x_1,...,x_n\right)\\ 0 & \text{otherwise}\end{Bmatrix}$$

$$\Rightarrow \max\left(x_1,...,x_n\right)-1\le\theta\le\min\left(x_1,...,x_n\right)$$

The χ^2 distribution with *n* degrees of freedom

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t}\, dt \qquad n > 0$$

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} \exp(-x/2) \qquad x > 0$$

$$E(x) = n \qquad \mathrm{var}(x) = 2n$$

There is a useful relation between the χ^2 and the normal distributions

Theorem connecting χ^2 and normal distributions

If the random variables $X_1,\ldots,X_n$ *are i.i.d. and* if each of these variables has standard normal distribution, then the sum of the squares

$$Y^2 = X_1^2 + \ldots + X_n^k$$

Has a χ^2 distribution with *n* degrees of freedom

The distribution functions

$$F(y) = \Pr(Y \le y) = \Pr\left(X^2 \le y\right) = \Pr\left(-y^{1/2} \le X \le y^{1/2}\right)$$

$$= \Phi\left(y^{1/2}\right) - \Phi\left(-y^{1/2}\right)$$

The p.d.f is obtained by differentiating both side $f(y) = F'(y)$

$\phi(y) = \Phi'(y)$. Note $\phi\left(y^{1/2}\right) = (2\pi)^{-1/2}\exp(-y/2)$. We have

$$f(y) = \phi\left(y^{1/2}\right)\left(1/2y^{-1/2}\right) + \phi\left(-y^{1/2}\right)\left(1/2y^{-1/2}\right)$$

$$f(y) = (2\pi)^{-1/2}\, y^{-1/2}\exp(-y/2)$$

which is the χ^2 distribution with one degree of freedom

Normal distribution: Parameters

Let $X_1,\ldots,X_n$ be a random sample from normal distribution having mean μ and variance σ^2 . Then the sample mean (hat denotes M.L.E)

$$\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{i=1,\ldots,n} X_i$$

and the sample variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1,\ldots,n} \left(X_i - \bar{X}_n\right)^2$$

are independent random variables.

$\hat{\mu}$ has a normal distribution with a mean μ and variance σ^2/n.

$n\hat{\sigma}^2 / \sigma^2$ has a chi-square distribution of *n-1* degrees of freedom

Why *n-1* ? (next slide)

Parameters of the normal distribution: Note 1

- Let $x_1, ..., x_n$ be a vector of random number of length *n* sampled from the normal distribtuion
- Let $y_1, ..., y_n$ be another vector of *n* random numbers, related to the previous vector by linear transformation *A* ($AA^t = I$)

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

- Consider now the calculation of the variance (next slide)

Variance

- The formula we should use for the variance

$$\mathrm{var}(X) = \frac{1}{n} \sum_{i=1,\ldots,n} (X_i - \mu)^2$$

- However, we do not know the exact mean, and therefore we use

$$\mathrm{var}(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

- What are the consequences of using this approximation?

Variance is not changing upon linear transformations

- Consider the expression

$$\sum_{i=1,\ldots,n} \left(Y_i - \bar{Y}_n\right)^2 = \sum_{i=1,\ldots,n} \left(\mathbf{A}X_i - \mathbf{A}\bar{X}_n\right)^t \left(\mathbf{A}X_i - \mathbf{A}\bar{X}_n\right)$$

$$\sum_{i=1,\ldots,n} \left(X_i^t - \bar{X}_n\right)^t \mathbf{A}^t\mathbf{A}\left(X_i^t - \bar{X}_n\right) = \sum_{i=1,\ldots,n} \left(X_i^t - \bar{X}_n\right)^2$$

- The analysis is based on the unitarity of **A**. Hence, linear transformation dos change the variance of he distribution. This makes it possible to exploit the difference between

$$\bar{X}_n \text{ and } \mu$$

The *n-1* (versus *n*) factor

- Since **A** is arbitrary (as long as it is unitary). We can choose one of the transformation vectors **a** to be $(1,\ldots,1)/n^{1/2}$
- The scalar product

$$X^t\mathbf{a} - \bar{X}_n\mathbf{a} = 0$$

- Is identically zero (remember how we compute the mean?)
- Hence since we computed the average from the same sample we computed the variance, the variance lost one degree of freedom.

The *n-1 factor II*

- Note that the *n-1* makes sense. Consider only a single sample point, which is of course very poor and leaves a high degree of uncertainty regarding the value sof the parameters. If we use *n* then the estimated variance becomes zero, while if we use *n-1* we obtain infinite, which is more appropriate to the problem at hand, for which we have no information to determine the variance

The *t* distribution (in preparation for confidence intervals)

- Consider two random variables *Y* and *Z*, such that *Y* has chi-2 distribution with n degrees of freedom and *Z* has a standard normal distribution the variable *X* is defined by

$$X = Z \Big/ \left(\frac{Y}{n^{1/2}} \right)$$

Then the distribution of *X* is the *t* distribution with *n* degrees of freedom.

The *t* distribution

- The function is tabulated and can be written in terms of Γ function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)\, dx$$

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{(n\pi)^{1/2}\,\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-(n+1)/2} \qquad \text{for } -\infty < x < \infty$$

- The *t* distribution is approaching the normal distribution as $n \to \infty$. It has the same mean but longer tails.

Confidence Interval

- Confidence interval provide an alternative to the use of estimator instead of the actual value of an unknown parameter. We can find an interval *(A,B) that we think has high probability of containing the desired parameter. The length of the interval gives us an idea how well we can estimate the parameter value.*

Confidence interval: Example

Sample distribution is normal with mean μ and standard deviation σ. We expect to find a sample *S* in the intervals

$$\mu_S \pm \sigma; \mu_S \pm 2\sigma; \mu_S \pm 3\sigma$$

About 68.27% 95.45% and 99.73 of the time respectively

Confidence interval for means

If the statistics *S* has the sample mean $\overline{X}$ then 95% and 99% confidence limits for estimation of the population mean are given by $\overline{X} \pm 1.96\sigma_{\overline{X}}$ and $\overline{X} \pm 2.58\sigma_{\overline{X}}$ respectively.

For large samples $(n \geq 30)$ we can write (depending on the level of confidence we are interested in) $\overline{X} \pm z_c \frac{\sigma}{\sqrt{n}}$

For small sample we need to *t* distribution

Confidence interval for the mean of the normal distribution

- Let $X_1, ..., X_n$ for a random sample from a normal distribution with unknown mean and unknown variance. Let $t_{n-1}(x)$ denote the p.d.f of the *t* distribution with *n-1* degrees of freedom, and let *c* be a constant such that

$$\int_{-c}^{c} t_{n-1}(x)\,dx = \gamma$$

- For every value of *n*, the value of *c* can be found from the table of the *t* distribution to fit the confidence (probability) γ

Confidence interval for means small sample (n<30)

- We use the special distribution *t* (instead the direct normal distribution expression) since the variance is estimated from the sample too. For 95% confidence in the mean

$$-t_{0.975} < \frac{\left(\bar{X} - \mu\right)\sqrt{n}}{\hat{\sigma}} < t_{0.975}$$

and the mean can be estimated as (with confidence 95%)

$$\bar{X} - t_{0.975}\frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{.975}\frac{\hat{\sigma}}{\sqrt{n}}$$