10 Apr 03
Neil E Cotter

Method of Moments:

Set estimated mean, $\hat{\mu} = \bar{x} \equiv \text{ave } x_i = \frac{1}{n}\sum_{i=1}^{n} x_i$

†Set estimated variance, $\hat{\sigma}^2 = s^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

†optional, use if needed

Now suppose $p(x)$ depends on parameters, (such as λ and k for gamma distribution), and these parameters are directly related to μ and σ^2.

Then we can estimate the parameters (such as k and λ for the gamma distribution) from $\hat{\mu}$ and $\hat{\sigma}^2$. This is the method of moments.

ex: The gamma distribution is $p(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$

for $x \geq 0$ where $k \equiv$ integer (# events)
$\lambda \equiv$ positive real (rate parameter)

We have $\mu = E\{X\} = \frac{k}{\lambda}$ and $\text{Var}\{X\} = \frac{k}{\lambda^2} = \sigma^2$.

$$\frac{\mu}{\sigma^2} = \frac{k/\lambda}{k/\lambda^2} = \lambda$$

$$\frac{\mu^2}{\sigma^2} = \frac{k^2/\lambda^2}{k/\lambda^2} = k$$

Thus, the method of moments says we estimate λ and k as follows: $\hat{\lambda} = \frac{\hat{\mu}}{\hat{\sigma}^2}$ and $\hat{k} = \frac{\hat{\mu}^2}{\hat{\sigma}^2}$

Comment: Usually, we invoke the Central Limit Theorem and argue that $p(x)$ is approximately gaussian. In that case we just use $\bar{x}$ and s^2 directly, and the method of moments says $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s^2$.

U U

ECE 3530 Parameter Estimates

10 Apr
Neil E

Maximum Likelihood Estimates:

Suppose we have samples x_i from $p(x)$.

Suppose also that $p(x)$ depends on parameters (such as μ and σ^2 if $p(x)$ is gaussian, $n(\mu, \sigma^2)$, K and λ if $p(x)$ is gamma distribution, or a and b if $p(x)$ is uniform distribution on interval $[a, b]$).

Assume the samples, x_i, are independent.

Then $p(x_1, \ldots, x_n) = p(x_1) \cdot \ldots \cdot p(x_n)$.

notn: $L(x_1, \ldots, x_n) \equiv p(x_1, \ldots, x_n)$ • "L" for likelihood

The maximum likehood estimate is to choose the parameters for $p(x)$ that maximize $L(x_1, \ldots, x_n)$.

Since $L(x_1, \ldots, x_n)$ is a product, we usually take $\ln L(x_1, \ldots, x_n)$ and find the maximum. This is allowed because $\ln()$ is a strictly increasing function and has its maximum at the same point as the argument has its maximum.

So we find $\max\limits_{\text{parameters}} \left[\ln L(x_1, \ldots, x_n) = \ln p(x_1) + \ldots + \ln p(x_n) \right]$.

μ, σ^2 if $p(x)$ gaussian

Note how taking the log of L changed multiplication into addition

From calculus, we get a maximum when a derivative is zero. So if we are finding μ and σ^2 for a gaussian, we would have:

$$\frac{\partial}{\partial \mu} \ln L(x_1, \ldots, x_n) = \frac{\partial \ln p(x_1)}{\partial \mu} + \ldots + \frac{\partial \ln p(x_n)}{\partial \mu} = 0$$

and $\dfrac{\partial}{\partial \sigma^2} \ln L(x_1, \ldots, x_n) = \dfrac{\partial}{\partial \sigma^2} \ln p(x_1) + \ldots + \dfrac{\partial}{\partial \sigma^2} \ln p(x_n) = 0$

For a gaussian, $\dfrac{\partial}{\partial \mu} \ln p(x) = \dfrac{\partial}{\partial \mu} \ln \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

$$= \frac{\partial}{\partial \mu}\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - (x-\mu)^2/2\sigma^2\right)$$

$$= \frac{\partial}{\partial \mu}\left[-(x-\mu)^2/2\sigma^2\right]$$

$$= -2(x-\mu)(-1)/2\sigma^2$$

$$= 2(x-\mu)/2\sigma^2$$

$$= (x-\mu)/\sigma^2$$

$$\frac{\partial}{\sigma^2} \ln p(x) = \frac{\partial}{\partial \sigma^2}\left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$= \sqrt{2\pi\sigma^2}\, \frac{\partial}{\partial \sigma^2} (2\pi\sigma^2)^{-\frac{1}{2}} + \frac{(x-\mu)^2}{2\sigma^4}$$

$$= (2\pi\sigma^2)^{\frac{1}{2}} \left(-\frac{1}{2}\right) (2\pi\sigma^2)^{-\frac{3}{2}} 2\pi + "$$

$$= -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}$$

Plugging in x_i and taking sum gives:

$$\sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma^2} = 0 \quad \text{and} \quad \sum_{i=1}^{n} -\frac{1}{2\sigma^2} + \frac{(x_i-\mu)^2}{2\sigma^4} = 0$$

Solving for μ and σ^2 (i.e., $\hat{\mu}$ and $\hat{\sigma}^2$) gives:

$$\hat{\mu} = \bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Note: We get the biased estimate for $\hat{\sigma}^2$ since we divide by n instead of $n-1$. That is,

$$E\{\hat{\sigma}^2\} = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

Estimates as Random Variables

The value of an estimated parameter such as $\hat{\mu}$ depends on random outcomes of experiments.

Consequently, $\hat{\mu}$ is the random outcome of an experiment (namely, taking the average of n random outcomes).

Thus, there is a probability density function, $p(\hat{\mu})$, for values of $\hat{\mu}$.

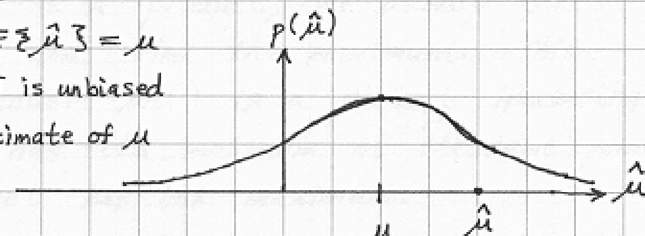
ex: Suppose the x_i are gaussian distributed and $\hat{\mu} = \bar{x} \equiv \frac{1}{n} \sum_{i=1}^{n} x_i$.

Then $\hat{\mu}$ is a sum of gaussian random outcomes, and $p(\hat{\mu})$ is a gaussian:

$$p(\hat{\mu}) \sim n(\mu, n \tfrac{1}{n^2} \sigma^2) = n(\mu, \tfrac{\sigma^2}{n})$$

$E\{\hat{\mu}\} = \mu$

$\Rightarrow \bar{x}$ is unbiased estimate of μ



We could get any $\hat{\mu}$ when we actually collect data, but we are more likely to get $\hat{\mu}$ values near μ.

We usually have to estimate σ^2, too. Our estimate $\hat{\sigma}^2$ also has an associated probability density function, $p(\hat{\sigma}^2)$.

ex: Suppose the x_i are gaussian distributed and $\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

ECE 3530 Parameter Estimates

10 Apr 03
Neil E Cotter

Then $\hat{\sigma}^2$ (we accept without proof) is distributed the same way as the sum of $n-1$ squared (independent) gaussian random variables. Thus, $\hat{\sigma}^2$ has a Chi-square distribution.

Recall: $X = X_1^2 + \ldots + X_n^2 \sim \chi_n^2$ Chi-square with n degrees of freedom.

$$p(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{n/2}\,\Gamma(n/2)}$$

We have $S^2 \sim \frac{\sigma^2}{n-1}\chi_{n-1}^2$ or $\frac{S}{\sigma} \sim \sqrt{\frac{\chi_{n-1}^2}{n-1}}$

$$\text{or } p(s^2) = \frac{\sigma^2}{n-1} \cdot \frac{s^{2\left(\frac{n-1}{2}-1\right)} e^{-\frac{s^2}{2}}}{2^{(n-1)/2}\,\Gamma\left(\frac{n-1}{2}\right)}$$

Note: We are assuming gaussian distributed x_i's.

We accept without proof that $\overline{X}$ and S^2 (the random variables associated with $\bar{x}$ and s^2) are independent.

Given the independence of $\overline{X}$ and S^2 plus the distribution for $\frac{S}{\sigma}$ that tells us how S and σ are related, we can determine the density function for $\frac{\overline{X}-\mu}{S}$. (Recall that this is like the term $\frac{X-\mu}{\sigma}$ used to transform a nonstandard gaussian to a standard gaussian.)

$$\text{We have } \frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\dfrac{\overline{X}-\mu}{\sigma}}{\dfrac{S/\sqrt{n}}{\sigma}} \sim \frac{n(0,1)/\sqrt{n}}{\sqrt{\dfrac{\chi_{n-1}^2}{n-1}}\,/\sqrt{n}} = t_{n-1}$$

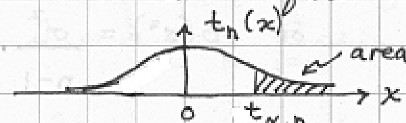
This is a t-distribution of degree n-1.

Student's or t-distribution (with n degrees of freedom)

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

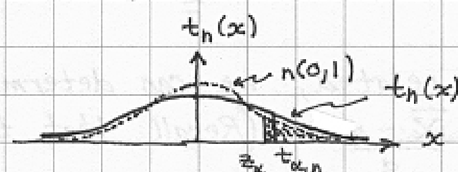
ref: "Probability and Mathematical Statistics, an Introduction," Eugene Lukacs, Academic Press, 1972.

Note: We are more interested in the <u>area</u> under the tails of $t_n(x)$ in practice. We use tables to determine the values of x that give certain values of areas under the tails. These areas are equal to integrals of $t_n(x)$.



Note: The t-distribution is symmetrical around $x = 0$.

Note: The t-distribution looks like a slightly flattened gaussian. It is slightly lower in the middle, and slightly higher on the sides.



Note: As $n \to \infty$, t_n approaches $n(0,1)$ standard gaussian.

The difference between the x value giving area under tail $= \alpha = 0.05$ for t_{30} and $n(0,1)$ is 1.697 vs 1.645 or about 3%. The discrepancy increases as α gets smaller. For $\alpha = 0.0005$ we get $t_{30} = 3.646$ vs $z_\alpha = 3.291$ or a difference of about 10%.