Math 152 2-10-09

Parameter Estimation

Setup:

observed data are regarded as instances of i.i.d r.v's

$$X_1, \ldots, X_n.$$

whose ~~joint~~ distribution depends on a parameter (vector) θ.

With the i.i.d assumption

$$f_{\vec{X}}(\vec{x}, \vec{\theta}) = f_X(x_1 | \theta) \cdot \ldots \cdot f_X(x_n | \theta).$$

An estimate for θ (called $\hat{\theta}$) will be a function of $X_1, \ldots, X_n$

$$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$$

and is a random variable.

We will be interested in the distribution of $\hat{\theta}$.

We will introduce two methods for estimating θ

1. Method of Moments
2. Method of Maximum Likelihood.

Method 1 is illustrative and sometimes easier to use.

Method 2 is more generally useful.

Method of Moments

$$\mu_k = E(X^k)$$

is the k^{th} moment of the distribution of the r.v. X.

If $X_1, \dots X_n$ are i.i.d from the same distribution as X

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

is the k^{th} sample moment.

Parameters which determine the distribution can often be expressed in terms of the moments.

e.g.

Poisson Distribution. $p(\lambda; k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$$\lambda = E(X).$$

$$\hat{\mu}_1 = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is the method of moments estimate for λ ; $\hat{\lambda} = \overline{X}$.

Note that $E(\hat{\lambda}) = E(\overline{X}) = \mu_1 = \lambda$.

so the estimate is <u>unbiased</u>.

What is the variability of $\hat{\lambda}$?

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Let $S = \sum_{i=1}^{n} X_i$

If X_i are independent Poisson R.V's with parameter λ_0 then

S is Poisson with parameter $(n \lambda_0)$.

<u>Recall</u>: (justifying the above statement).

$X \sim P(\lambda, k)$

$$E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t}$$

$$= e^{\lambda(e^t - 1)}.$$

If $X_1, \ldots, X_n$ are ind $P(\lambda, k)$ then

$$E\left(e^{t(X_1 + \cdots + X_n)}\right) = E(e^{tX_1}) \cdots E(e^{tX_n}) = e^{(n\lambda)(e^t - 1)}.$$

$\therefore$ $X_1 + \cdots + X_n$ is Poisson w/ parameter $n\lambda$.

recall also that

$$E(X^2) = \left(\frac{d}{dt}\right)^2 \left(e^{\lambda(e^t-1)}\right)\Big|_{t=0}.$$

$$\frac{d}{dt}\left(e^{\lambda(e^t-1)}\right) = \lambda e^t e^{\lambda(e^t-1)} \quad \left(E(X)=\lambda\right)$$

$$\left(\frac{d}{dt}\right)^2\left(e^{\lambda(e^t-1)}\right) = \lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)}$$

$$E(X^2) = \lambda + \lambda^2.$$

$$Var(X) = E(X^2) - E(X^2) = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

We have then

$$E(\hat{\lambda}) = \frac{1}{n} E(S) = \lambda_0.$$

$$Var(\hat{\lambda}) = \frac{1}{n^2} Var(S) - \frac{\lambda_0}{n}.$$

We can bring the Central limit theorem into the discussion...

Consider $S^* = \dfrac{S - n\lambda_0}{\sqrt{n\lambda_0}}$.

$$E(e^{tS^*}) = e^{-t\cdot\sqrt{n\lambda_0}} \cdot E\left(e^{\frac{t}{\sqrt{n\lambda_0}} S^*}\right)$$

$$= e^{-t\sqrt{n\lambda_0}} \; e^{n\lambda_0\left(e^{t/\sqrt{n\lambda_0}} - 1\right)}$$

One shows that this $\longrightarrow e^{t^2/2}$

as $n \to \infty$.

and this ($e^{t^2/2}$) is the m.g.f. for a standard normal.

Consequently, for large $n\lambda_0$, the distribution of S is approximately normal.

$\therefore$ The distribution of $\hat{\lambda}$ is approximately normal with mean λ_0 and $\operatorname{Var}(\hat{\lambda}) = \dfrac{\lambda_0}{n}$

The standard error of $\hat{\lambda}$

is $\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda_0}{n}}$.

As usual (from last chapter) we don't know λ_0 and for practical purposes we work with the estimated standard error

$$s_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{n}}.$$

e.g. asbestos fibers

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

$\hat{\lambda} = 24.9$. $\quad s_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{23}} \approx 1.04$

If the model is correct

95% / $(24.9 - (1.96)1.04,\ 24.9 + (1.96)(1.04))$.

How do we know if the model is correct?
We will take up that question in Chapter 9.

e.g. Normal Distribution (Method of Moments).

$$X \sim N(\mu, \sigma^2).$$

$$\mu_1 = E(X) = \mu.$$

$$\mu_2 = E(X^2) = \mu^2 + \sigma^2$$

$$\mu = \mu_1$$

$$\sigma^2 = \mu_2 - \mu_1^2.$$

The estimates of μ and σ^2 from the sample moments are.

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Recall: if $X \sim N(0,1)$.

$$E(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} \, dx.$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2tx + t^2)} \; e^{t^2/2} dx$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} \, dx.$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 u^2} du. = e^{t^2/2}.$$

and $\sigma X + \mu$ is $N(\mu, \sigma^2)$.

with $E(e^{t(\sigma X + \mu)}) = e^{\mu t} e^{\sigma^2 t^2/2}$

if $X_1, \ldots X_n$ are i.i.d. $N(\mu, \sigma^2)$

then $E(e^{t(X_1 + \ldots + X_n)}) = e^{n\mu t} e^{(\sqrt{n}\sigma)^2 t^2/2}$

So $X_1 + \ldots + X_n$ is $N(n\mu, n\sigma^2)$

and $\frac{1}{n}(X_1 + \ldots + X_n)$ is $N(\mu, \frac{\sigma^2}{n})$.

So we know the sample distribution of $\hat{\mu} = \bar{X}$ $(N(\mu, \sigma^2/n))$.

and one can show (see chpt 6) that

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$$

e.g. Gamma Distribution (method of moments)

$X \sim \text{gamma}(\alpha, \lambda)$

$$g_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} \qquad t \geq 0.$$

$$= 0 \qquad t < 0.$$

where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du. \quad x > 0.$

$$E(e^{tX}) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x} dx$$

For any α, λ

$$\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}\, dt = \underline{1}.$$

So $$\frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\lambda-t)}\, dx = 1.$$

and we have.

$$E(e^{tX}) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$$

$$\frac{d}{dt}\left(\frac{\lambda}{\lambda-t}\right)^\alpha = (\lambda^\alpha)\left(-\alpha(\lambda-t)^{-\alpha-1}\right)(-1)$$

$$= \frac{\lambda^\alpha \cdot \alpha}{(\lambda-t)^{(\alpha+1)}}$$

$$\left(\frac{d}{dt}\right)^2\left[\left(\frac{\lambda}{\lambda-t}\right)^\alpha\right] = \lambda^\alpha \cdot \alpha\left(-(\alpha+1)(\lambda-t)^{-(\alpha+2)} \cdot (-1)\right)$$

$$= \frac{\lambda^\alpha \cdot \alpha \cdot (\alpha+1)}{\cancel{(\alpha+1)}(\lambda-t)^{\alpha+2}}$$

$$E(X) = \frac{\alpha}{\lambda} \qquad E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

So $\operatorname{Var}(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2$

$= \frac{\alpha}{\lambda^2}.$

We have (back to method of moments).

$\mu_1 = \frac{\alpha}{\lambda}$

$\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}.$

Solving for α, λ:

$\alpha = \mu_1 \lambda.$

$\mu_2 = \frac{\mu_1 \lambda (\mu_1 \lambda + 1)}{\lambda^2}.$

$= \frac{\mu_1^2 \lambda^2 + \mu_1 \lambda}{\lambda^2} = \cancel{\mu_1^2 \lambda + \mu_1} \quad \mu_1^2 + \frac{\mu_1}{\lambda}.$

$\cancel{\frac{\mu_2 - \mu_1}{\mu_1^2}} \quad \frac{\mu_2 - \mu_1^2}{\mu_1} = \frac{1}{\lambda}$ $\quad \boxed{\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}.}$

$$\boxed{\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.}$$

$$\hat{\lambda} = \frac{\bar{X}}{\hat{\sigma}^2} \quad \text{since } \hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2.$$

$$\hat{\alpha} = \frac{(\bar{X})^2}{\hat{\sigma}^2}$$

In the previous two examples it was feasible to find ~~an analytic~~ explicit expressions for the sample distributions of the estimated parameters. but in the Gamma distribution case it is not feasible.

To assess the variability of the estimates $\hat{\lambda}$ and $\hat{\alpha}$ we can use the Bootstrap technique.

Suppose we knew the "true" values of α and λ (α_0 and λ_0).
and our sample has size n.
On a computer, generate a large number (M) of samples of size n from a $\Gamma(\alpha_0, \lambda_0)$ distribution.

$$\vec{S}_i = \{ X_{1i} , X_{2i} \dots X_{ni} \}$$

$$i = 1, \dots M.$$

Use $\vec{S}_i$ to determine estimates

$\hat{\alpha}_i^*$ and $\hat{\lambda}_i^*$

Then normalized histograms of the

$\hat{\alpha}_i^*$ $i=1,\dots M$ and $\hat{\lambda}_i^*$ $i=1,\dots M$.

give approximations to sample distributions of $\hat{\alpha}, \hat{\lambda}$.

Since we don't know the "true" values we use what is available: our original sample values of $\hat{\alpha}, \hat{\lambda}$.

i.e. we generate many (M) samples of size n from a gamma$(\hat{\alpha}, \hat{\lambda})$ distribution and use each sample to give an estimate $\hat{\alpha}_i^*, \hat{\lambda}_i^*$ $i = (1, \ldots M)$.

It can be shown that in many practical situations the variability of the $\hat{\alpha}_i^*$ and $\hat{\lambda}_i^*$ will be a good assessment of the variability of our original $\hat{\alpha}, \hat{\lambda}$.