Contributeurs

Note The moments of a mixture, are the mixture of the moments.

Lesson 25 : Estimator Quality

Sample Statistics

Sample Mean Unbiased estimator of the true mean μ.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sample Variance Unbiased estimator of the true variance σ^2.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

Empirical Variance Biased estimator of the true variance σ^2.

$$\hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

Lesson 27 : Method of Moments

Notation

μ'_k k^{th} moment centred around 0, $\boxed{\mu'_k = \mathrm{E}[X^k]}$.

$\hat{=}$ Set equal.

Exponential Distribution There is only one parameter θ which is the mean, we set $\boxed{\hat{\theta} \mathrel{\hat{=}} \mu'_1}$.

Gamma Distribution We have :

$$\mathrm{E}[X] = \alpha\theta \mathrel{\hat{=}} \bar{x} \qquad \mathrm{Var}(X) = \alpha\theta^2 \mathrel{\hat{=}} \hat{\sigma}^2$$

$$\therefore \hat{\theta} = \frac{\hat{\sigma}^2}{\bar{x}} = \frac{\hat{\mu}'_2 - \hat{\mu}'^2_1}{\hat{\mu}'_1} \qquad \hat{\alpha} = \frac{\bar{x}^2}{\hat{\sigma}^2} = \frac{\hat{\mu}'^2_1}{\hat{\mu}'_2 - \hat{\mu}'^2_1}$$

Pareto Distribution

$$\mathrm{E}[X] = \frac{\theta}{\alpha - 1} \mathrel{\hat{=}} \hat{\mu}'_1 \qquad \mathrm{E}\left[X^2\right] = \frac{2\theta^2}{(\alpha-1)(\alpha-2)} \mathrel{\hat{=}} \hat{\mu}'_2$$

$$\therefore \hat{\alpha} = \frac{2(\hat{\mu}'_2 - \hat{\mu}'^2_1)}{(\hat{\mu}'_2 - 2\hat{\mu}'^2_1)} \qquad \hat{\theta} = \frac{\hat{\mu}'_1\hat{\mu}'_2}{\hat{\mu}'_2 - 2\hat{\mu}'^2_1}$$

Lognormal Distribution

$$\hat{\mu} = 2\ln(\hat{\mu}'_1) - 0.5\ln(\hat{\mu}'_2) \qquad \hat{\sigma}^2 = \ln(\hat{\mu}'_2) - 2\ln(\hat{\mu}'_1)$$

Uniform Distribution

$$\mathrm{E}[X] = \frac{\theta}{2} \qquad \therefore \hat{\theta} = 2\hat{\mu}'_1$$

Lesson 28 : Percentile Matching

Note : Exams don't typically ask a lot of percentile matching questions, thus it's not really worth memorizing each distribution's formulas as it can easily be about a random distribution.

Exponential Distribution

$$\hat{\theta} = \frac{-\pi_g}{\ln(1-g)}$$

Weibull Distribution

$$\hat{\tau} = \frac{\ln\left(\ln(1-g_1)\,/\ln(1-g_2)\right)}{\ln(\pi_{g_1}/\pi_{g_2})} \qquad \hat{\theta} = \frac{\pi_{g_1}}{\sqrt[\hat{\tau}]{-\ln(1-g_1)}}$$

Lognormal Distribution (use the percentiles of a normal distribution z_p)

$$\hat{\sigma} = \frac{\ln(\pi_{g_2}) - \ln(\pi_{g_1})}{z_{g_2} - z_{g_1}} \qquad \hat{\mu} = \ln(\pi_{g_1}) - z_{g_1}\hat{\sigma}$$

Truncated data

For $(X|X > d)$:

$$F_X(x|X > d) = \frac{F_X(x) - F_x(d)}{S_X(d)} \qquad S_X(x|X > d) = \frac{S_X(x)}{S_X(d)}$$

Lesson 29 :

Lesson 30 : MLE Special Techniques

If the likelihood function is of the form $\boxed{\mathcal{L}(\gamma) = \gamma^{-a}\mathrm{e}^{-b/\gamma}}$ then $\boxed{\hat{\gamma}^{\text{MLE}} = \frac{b}{a}}$.

If the likelihood function is of the form $\boxed{\mathcal{L}(\lambda) = \lambda^{a}\mathrm{e}^{-\lambda b}}$ then $\boxed{\hat{\lambda}^{\text{MLE}} = \frac{a}{b}}$.

If the likelihood function is of the form $\boxed{\mathcal{L}(\theta) = \theta^{a}(1-\theta)^{b}}$ then $\boxed{\hat{\theta}^{\text{MLE}} = \frac{a}{a+b}}$.

Exponential

$$\hat{\theta} = \frac{\sum_{i=1}^{n+c}(x_i - d_i)}{n}$$

Weibull with a fixed τ

$$\hat{\theta} = \sqrt[\tau]{\frac{\sum_{i=1}^{n+c} x_i^\tau - \sum_{i=1}^{n+c} d_i^\tau}{n}}$$

Lognormal (use the percentiles of a normal distribution z_p)

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \ln^2 x_i}{n} - \hat{\mu}^2} \qquad \hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

Uniform $(0, \theta)$ for individual data

$$\hat{\theta} = \max x_i$$

Uniform $(0, \theta)$ for grouped data

c_j Upper bound of highest finite interval

n_j Number of observations below c_j

$$\hat{\theta} = c_j \left(\frac{n}{n_j} \right)$$

Inverse exponential

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} (1/x_i)}$$

Two-parameter Pareto, fixed θ

$$\hat{\alpha} = -\frac{n}{K} \qquad K = \sum_{i=1}^{n+c} \{\ln(\theta + d_i) - \ln(\theta + x_i)\}$$

Single-parameter Pareto, fixed θ

$$\hat{\alpha} = -\frac{n}{K} \qquad K = \sum_{i=1}^{n+c} \{\ln \max(\theta, d_i) - \ln x_i\}$$

Beta, fixed θ, $\beta = 1$

$$\hat{\alpha} = -\frac{n}{K} \qquad K = \sum_{i=1}^{n+c} \{\ln(x_i)\} - n \ln \theta$$

Beta, fixed θ, $\alpha = 1$

$$\hat{\beta} = -\frac{n}{K} \qquad K = \sum_{i=1}^{n+c} \{\ln(\theta - x_i)\} - n \ln \theta$$

Lesson 31 :

Lesson 51 : Time Series : Trend and Seasonality

Notation

At time t,

$\{x_t\}$ Observed time series.

- Formally, a time series of length n is written as $\{x_t : t = 1, 2, \ldots, n\} = \{x_1, x_2, \ldots, x_n\}$.
- For simplicity, we write x_t.

m_t Trend.

s_t Seasonal effect.

z_t Error term.

- It is generally a sequence of correlated variables with mean zero.

Additive decomposition model

$$x_t = m_t + s_t + z_t$$

Estimation of seasonal variation :

$$\hat{s}_t = x_t - \hat{m}_t$$

Multiplicative model

When the seasonal effect s_t tends to increase as the trend m_t increases (i.e., $s_t \propto m_t$) :

$$x_t = m_t \cdot s_t + z_t$$

When the trend itself is multiplicative :

$$x_t = m_t \cdot s_t \cdot z_t$$

or

$$\ln(x_t) = \ln(m_t) + \ln(s_t) + \ln(z_t)$$

- However, The logarithmic model must be used with caution ; If z_t is normally distributed then $\ln(z_t)$ is lognormally distributed and thus has a *greater mean*.

Estimation of seasonal variation :

$$\hat{s}_t = \frac{x_t}{\hat{m}_t}$$

Calculating seasonal variation

For example, we suppose monthly data.
So,

1. We average the seasonal variations $\hat{s}_t$ for each month.
2. For an additive model, we then subtract from the respective month's seasonal variations these twelve averages.
3. For a multiplicative model, we then divide the respective month's seasonal variations by these twelve averages.

This leads to an average seasonal variation of 0 for each month.

Center moving average for monthly data :

$$\hat{m}_t = \frac{0.5m_{t-6} + \sum_{i=t-5}^{t+5} m_i + 0.5m_{t+6}}{12}$$