Boston University Questrom School of Business

MF793 – Fall 2021

Eric Jacquier

Properties of Estimators and Convergence Results
Normal and Lognormal Distributions

- An estimator is a function of the data used to estimate an unknown parameter.

- What properties do typical estimators have?

- What properties would we like them to have?

- Convergence to the normal distribution.

- Summaries of properties of the normal, univariate and bivariate, and lognormal distributions

Why learning classical statistics is hard for smart people

1. Stats 101 Prof: *"The true, unknown parameter is a fixed number"*

(Normally) smart student: strange since nobody will ever know the parameter anyways.

Philosopher: Good point, our brain works so we think of the unknown as uncertain, …, random.

2. Stats 101 Prof: *"The **estimator** is random, because the data are random."*

Student: What do you mean, the data are random ? I just got the data, they are known numbers. I just got an estimate of the mean, I found 0.156, that's not random?

Prof: 0.156 is the **estimate** for this dataset, not the **estimator**.
The ***estimator*** represents the randomness of the ***estimate*** in repeated sampling
If you had drawn another random dataset, you would have another estimate.

Student: It gets weirder. I must worry about what the estimate, which I already have, *could have been* if I had gotten *data I did not get* ?
Can't we just try to see what we learnt given the data we got?

Prof: The unknown parameter is a fixed number and the estimator is a random variable.
If you find this strange, you can learn Bayesian statistics later.

- Hopefully as the number of observations grows, the **estimator** get more precise and “closer” to the true, fixed, unknown parameter being estimated.

- For any estima**tor** we need

  1. Its expected value (is the estima**tor** *unbiased* over repeated samples?)

  2. Some measure of precision of the estimator (variance, mean squared error)

  3. Its distribution

- Keep in mind, we often want to estimate functions of the parameters

  Portfolio R~N(μ, σ)   Sharpe Ratio = (μ - R_f) / σ

- Various convergence results help us

$$E\left[\frac{\bar{x}}{\bar{y}}\right] = \frac{E(\bar{x})}{E(\bar{y})}$$

Convergence in Probability (weak convergence) and consistency

- <u>Definition</u>: A sequence of random variables X_1, X_2, ... X_n, converges in probability to a random variable X if, for every $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \qquad [1]$$

  - X doesn't need to be random, it can be a fixed number,
(.... like a true unkown parameter !)

  - X_n can be a sample mean with n observations or any parameter estimator. …then the "sequence" in [1] is for increasing sample sizes

- **Weak Law of Large numbers**:

  Prove that [1] is true for the sample mean with X=μ: $\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$

The sample mean converges in probability to the true mean μ

Proof: apply Chebyshev inequality: $P((\bar{X}_n - \mu)^2 > \varepsilon^2) < \frac{Var(\bar{X}_n)}{\varepsilon^2} =$

- This property is known as **consistency**:

  "The estimator approaches the true parameter as the sample grows larger"

- An estimator can be biased (in small sample) but consistent ! E.g.,

$\frac{1}{T-1}\Sigma(R_t-\bar{R})^2$

$$\widehat{\sigma^2} = \frac{1}{T}\sum_{1}^{T}(R_t - \bar{R})^2$$

- **How about Functions of random variables ?**

We often need to estimate *functions of* the parameters:

Examples: Sharpe ratio: $S_h = (\mu - R_f) / \sigma$ $\widehat{S_h} = ?$ $\frac{\hat{\mu}-R_f}{\hat{\sigma}}$

  - Given consistent estimators for each parameter μ, σ:

    Is a function of the estimators a consistent estimator of the function of the parameter?

  - That is: Can we just "***plug in***" estimates in the function and hope for consistency?

Result: **If X_n converges in probability to X, and f is a continuous function of X, Then $f(X_n)$ converges in probability to f(X)**

  - Notion of ***plim*** (probability limit): plim(f($\hat{\sigma}$, $\hat{\mu}$)) = f (plim($\hat{\sigma}$) , plim($\hat{\mu}$))

  - Recall that it is not true in expectation in small sample – due to the Jensen effect.
  - In **large sample** the Jensen effect vanishes

Distribution of the sample mean

- Recall, for an i.i.d. sample of N observations $X_i \sim (\mu, \sigma^2)$, the sample mean $\bar{X} \sim (\mu, \frac{\sigma^2}{N})$

- **Central Limit Theorem:**

  If the random variables $X_1, \ldots, X_n$ form a random sample of size n from a distribution with mean μ and **finite variance σ^2**, then for any fixed number x_0.

$$\lim_{n \to \infty} \Pr\left[\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} < x_0\right] = \Phi(x_0)$$

$$\sum_{1}^{T} \frac{(X_t - \bar{X})^2}{T-1}$$

  where Φ is the normal CDF.

  We already knew the mean and variance of $\overline{X_n}$, we now know its distribution, approximately, for large samples. ... ***asymptotically***

- The CLT can be extended to **independent but not necessarily identically** distributed samples. Again, the asymptotic normality of the sample mean follows. (Liapounov)

- If the sample itself is normally distributed, we don't need a central limit theorem: The mean is normally distributed even for a small sample as an average of normal RVs.

- The CLT also implies asymptotic normality of estimators of other quantities, .i.e., σ^2, K, Sk., etc., because they are just sample means of different RVs

Properties of estimators

- **Consistency**: an estimator $\widehat{\theta_n}$ is a consistent estimator of θ if for every ε>0 and every θ

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| > \varepsilon\right) = 0$$

- **Bias / unbiasedness**: are small sample concepts, no convergence to large sample is needed.

  Any bias vanishes in large sample if the estimator is consistent

  An estimator can be biased, yet consistent and even asymptotically efficient:

- **Asymptotic efficiency**:

  An estimator is asymptotically efficient if it is consistent and no other estimator has a smaller asymptotic variance.

- **Mean Squared error (MSE)** is a small sample concept: Prove that

$$E\left[(\hat{\theta} - \theta)^2\right] = \left[E(\hat{\theta} - \theta)\right]^2 + Var(\hat{\theta}) = E\left[\left[\hat{\theta} - E(\hat{\theta})\right]^2\right]$$

$$\underbrace{E\left[(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^2\right]}_{MSE} = \left[E\hat{\theta} - \theta\right]^2 + E\left[(\hat{\theta} - E\hat{\theta})^2\right] + 2\underbrace{E[(\ )(\ )]}_{=0}$$

$$= BIAS^2 + Var$$

Normal Distribution

- Fundamental building block: the Normal arises from the averaging of other distributions under general conditions (variance must exist)

<u>Definition:</u> **Normal PDF:**

$$f(x|\mu,\sigma^2) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \text{for } -\infty < x < \infty.$$

- The Normal <u>quickly</u> rules out extreme outcomes: Prob(|x-μ|>kσ) gets small quickly as k increases

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
5	$1 - 6 \times 10^{-7}$

- The normal is preserved by addition: Linear combinations of normal RVs are normal.

- The normal density is maximum entropy given $X \in (-\infty, \infty)$, and a known mean / variance.

Bivariate Normal

You need to be able to prove everything on this page

Definition: Two RVs (X,Y) are bivariate normal with correlation ρ if

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\} \quad [1]$$

- ρ=0 => p(x,y) = n(x) n(y) … Makes sense

$= \hat{A} + \hat{B}(X - \mu_x)$

- We can rewrite p(x,y) = N(y|x) N(x) with: $E(Y|X) = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)$ [2]

  (Prove it) $e^{-\frac{1}{2}\frac{(y-E[y|x])^2}{V(y|x)}}$ and: $V(Y|X) = \sigma_y^2(1-\rho^2)$ [3]

  $V(Y)$

- [2]: Mean of Y given X: is a linear function of X, the ***regression*** of Y on X

- [3]: Variance of Y given X: the variance of the ***error*** of the regression,
  … the part of V(Y) not explained by X: V(Y|X)

- $\frac{\rho\sigma_y}{\sigma_x} = \frac{\rho\sigma_y\sigma_x}{\sigma_x^2} = \frac{\sigma_{xy}}{\sigma_x^2} = \boldsymbol{\beta}$ The **beta** of y with respect to x, The slope of the regression.

Lognormal distribution

You need to be able to prove every result on this page

- We often model stock returns as log-normal rather than normal, namely:

$$r = \text{Log}(1+R) \sim N(\mu, \sigma)$$

<u>Definition:</u> **$Y = e^X$ is said to be lognormally distributed if $X = \log(Y) \sim N(\mu,\sigma)$.**

- Find the lognormal pdf by the change of variable method. $X = \log Y$ $\quad \frac{dx}{dy} = \frac{1}{y}$

$$p_Y(Y) = p_X(\log(Y)) \; | df^{-1} X / dY |$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left\{-\frac{1}{2\sigma^2}(\log Y - \mu)^2\right\} \cdot \frac{1}{y}$$

- Mean: $E(Y) = e^{\mu + 0.5\,\sigma^2}$ write $Y = e^{\mu+\sigma z}$

$$\int e^{\mu+\sigma z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z^2)\right\} = e^{\mu} \int \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}(z^2 - 2\sigma z + \sigma^2) + \frac{\sigma^2}{2}\right)}$$

$$= e^{\mu + \frac{1}{2}\sigma^2} \underbrace{\int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\sigma)^2} dz}_{1}$$

- Median: $M(Y) = e^{\mu}$

$Pr(e^x < e^{x_0}) = q_0 = .10$

- $E(Y^2) =$

$y^2 = e^{2\mu + 2\sigma z}$

- $V(Y) = E(Y^2) - E(Y)^2 = e^{2\mu + \sigma^2}[e^{\sigma^2} - 1]$ No need to learn if you know the pieces