Prob & Stat Review

February 19, 2019

Conditional Prob

- For two Events A, B, $s.t.\ Pr(B) > 0$,
  $$P(A|B) = \frac{P(A\ \ and\ \ B)}{P(B)}$$
  $$\Rightarrow P(AB) = P(A|B)P(B)$$
- $P(ABC) = P(AB|C)P(C) = P(A|BC)P(B|C)P(C)$
- A and B are independent if $P(AB) = P(A)P(B)$

Random Variable

- $X \overset{\text{pdf}}{\sim} f_x(\cdot) \quad \Leftrightarrow \quad Pr(x \in A) = \int_A f_X(x)\mathrm{d}x$
- $\forall x \in R, f_X(x) \geqslant 0$ and $\int f_X(x)\mathrm{d}x = 1$
- CDF : $\quad F_X(x) = Pr(X \leqslant x) = \int_{-x}^{x} f_X(s)\mathrm{d}s$
- Normal pdf. $\quad \Phi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\{ -\frac{(x-\mu)^2}{2\sigma^2} \}$

Expectation / Variance / Moment

$$\mu_k(X) = \int (x-\mu)^k f_X(x)\mathrm{d}x.$$

- Mean: $k = 1, \quad E(X) = \mu_1$
- Variance:
  $\sigma_X^2 = Var(X) = \mu_2 = \int (x-\mu)^2 f_X(x)\mathrm{d}x = E(X-\mu)^2$
- Skewness: $\mu_3 = E(X-\mu)^3$
- $\mu_4 = E(X-\mu)^4$
- $m_k(X) = \int x^k f_X(x)\mathrm{d}x$

Multivariate Distribution

- Joint density: $(X, Y) \sim f_{X,Y}(\,\cdot\,,\,\cdot\,)$
- $Pr\Big((X, Y) \in A\Big) = \int\int_A f_{XY}(x, y)\mathrm{d}x\mathrm{d}y.$
- Conditional density: $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$
- Marginal density: $f_X(x) = \int f_{X,Y}(x, y)\mathrm{d}y$
- Conditional expectation: $E(Y|X = x) = \int y f_{Y|X}(y|x)\mathrm{d}y$
- Conditional variance:
  $Var(Y|X = x) = \int (y - E(Y|X = x))^2 f_{Y|X}(y|x)\mathrm{d}y$
- Independence: iff $\underline{f_{X,Y}(x, y) = f_X(x)f_Y(y)}$, X and Y are independent.

Transformation of R.V.

$Y = r(X)$ and $X \overset{\text{pdf}}{\sim} f_X$,

- the *cdf* of Y is

$$G(y) = Pr(Y \leqslant y) = Pr(r(X) \leqslant y) = \int_{\{x:\ r(X) \leqslant y\}} f(x)\mathrm{d}x,$$

- if r is monotonic, the *pdf* of Y is

$$g(y) = f(s(y))\Big|\frac{\mathrm{d}s(y)}{\mathrm{d}y}\Big|,$$

where $s(\cdot)$ is the inverse function of r. i.e. $x = s(y)$.

Properties of Expectation & Variance, Covariance

- Linearity: $E(aX + bY + c) = aE(X) + bE(Y) + c$
- $E(XY) \xlongequal{X \ \& \ Y independent} E(X)E(Y)$
- X and Y independent $\Leftrightarrow E\left[g(X)h(Y)\right] = E\left[g(X)\right]E\left[h(Y)\right]$ for any measurable g, h.
- $Var(aX + b) = a^2 Var(X)$
- $Var(X) = E(X^2) - (EX)^2$
- if X and Y are independent, $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$
- Covariance: $Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right] = E(XY) - E(X)E(Y)$
- Correlation: $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \ \in [-1, 1]$

Properties of Expectation & Variance, Covariance

- $Cov(X, Y) = \rho(X, Y) = 0$ if X and Y are independent.
- if $Cov(X, Y) = 0$, X and Y are said to be uncorrelated.
- $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$
- $Cov(aX, bY) = abCov(X, Y)$
- $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2\sum_{i<j}\sum Cov(X_i, X_j) \overset{if\ X_1,\ldots,X_n\ uncorrelated}{=\!=\!=\!=\!=\!=} \sum_{i=1}^n Var(X_i)$
- Law of iterated expectation: $E(E(Y|X)) = E(Y)$
- Eve's Law: $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$

Random Sample

$\{X_1, \dots, X_n\}$ is called a random sample if each X_i has been taken from the same distribution and is independently from each other (i.i.d).

Suppose $E(X_i) = \mu, Var(X_i) = \sigma^2$

- Sample Mean: $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$
- $E(\overline{X}_n) = E(X) = \mu$
- $Var(\overline{X}_n) = Var(\frac{1}{n} \sum X_i) = \frac{1}{n^2} \sum Var(X_i) = \frac{\sigma^2}{n}$

LLN and CLT

- Converges in Prob:
  $Z_n \xrightarrow{P} b \Leftrightarrow \lim_{n\to\infty} Pr(|Z_n - b| < \varepsilon) = 1$
- LLN: $\underline{\overline{X}_n \xrightarrow{P} \mu}$ (Chebyshev inequality)
- $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ if $X_i \sim N(\mu, \sigma^2)$
- CLT: if $X_i \sim F(\mu, \sigma^2)$, $\overline{X}_n \overset{A}{\sim} N(\mu, \frac{\sigma^2}{n})$
  or $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$

Estimator & MLE

$\{X_1, \ldots, X_n\}$ is a random sample and $X_i \sim f_X(\,\cdot\,; \theta_0)$,

- $L_n(\theta) \overset{Def.}{=} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) \overset{i.i.d}{=} \prod_{i=1}^{n} f_X(x_i; \theta)$
- $\hat{\theta}_{MLE} = argmax_{\{\theta \in \Theta\}} L_n(\theta)$

Under a general set of conditions,

- $\hat{\theta}_{MLE} \xrightarrow{P} \theta_0$
- $\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{d} N\big(0, I^{-1}(\theta)\big)$ where $I^{-1}(\theta)$ is the information matrix.

Estimator & MLE

- Statistic: $\hat{\theta} = \delta(X_1, \ldots, X_n)$
- Unbiasedness: if $E_\theta \hat{\theta} = \theta$ for any $\theta \in \Theta$
- Consistency: $\hat{\theta} \xrightarrow{P} \theta_0$
- $MSE(\hat{\theta}; \theta_0) = E(\hat{\theta} - \theta_0)^2 = Var(\hat{\theta}) + \left[bias(\hat{\theta})\right]^2$
- $bias(\hat{\theta}) = E(\hat{\theta} - \theta_0)$

Hypothesis Testing

Remark 1: $X \sim N(0,1)$ and $Y \sim N(0,1)$, then $X^2 \sim \chi^2(1)$, and if $X \perp Y$, $X^2 + Y^2 \sim \chi^2(2)$.

- $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2, \quad \sum_{i=1}^n (X_i - \overline{X}_n)^2/\sigma^2 \sim \chi^2_{n-1}$

Remark 2: $Z \sim N(0,1)$, $Y \sim \chi^2_n$, and $Z \perp Y$, then $X = \frac{Z}{\sqrt{Y/n}} \sim t_n$.

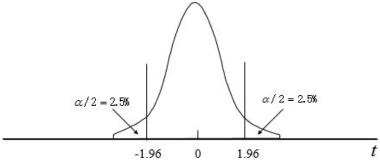
Example

- $Z = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1)$
- $t = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S} \sim t_{n-1}$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$

Remark 3: $Y \sim \chi^2_m$, $W \sim \chi^2_n$, and $Y \perp W$, then $F = \frac{Y/m}{W/n} = \frac{nY}{mW} \sim F_{m,n}$.

Hypothesis Testing

Confidence Interval, Hypothesis testing, p-value



$$-1.96 \ < \ \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma} \ < \ 1.96$$
$$\Rightarrow Pr\{\mu \in [\overline{X}_n - 1.96 * \sigma/\sqrt{n}, \overline{X}_n + 1.96 * \sigma/\sqrt{n}]\} = 95\%$$

$$H_0 : \quad \mu = \mu_0 \quad v.s. \quad H_1 : \quad \mu \neq \mu_0$$

if $\left|\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}\right| > 1.96$, reject H_0 at 5% sig. lev.
or equivalently reject H_0 if
P-value$= Pr(|z| > \left|\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}\right|) < 0.05$ (any given sig. level)

Regression & LS

$$Y_i = X_i'\beta + \varepsilon_i \;\; = X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{ik}\beta_k + \varepsilon_i, \quad Var(\varepsilon_i) = \sigma^2.$$

Vector Form: $Y = X\beta + \varepsilon$

- $\hat{\beta} = (X^TX)^{-1}X^TY = argmin \; \sum_{i=1}^{n}(Y_i - X_i'\beta)^2 \; = argmin \; (Y - X\beta)^T(Y - X\beta)$
- $\hat{\sigma}^2 = (Y - X\hat{\beta})^T(Y - X\hat{\beta}) \; / \; (n-k)$
- $\hat{\beta}_k \sim N\big(\beta_k, \sigma^2[(X^TX)^{-1}]_{kk}\big)$
- $H_0 : \;\; \beta_k = \overline{\beta}_k, \quad t_{\beta_k} = \frac{\hat{\beta}_k - \overline{\beta}_k}{\sqrt{\hat{\sigma}^2[(X^TX)^{-1}]_{kk}}}$
- $E(\hat{\beta}) = \beta$ if $E(\varepsilon|X) = 0$
  $\Rightarrow \;\; E(\varepsilon X) = 0, \; i.e. \; E(\varepsilon_i X_k) = 0$ (Strict Exogeneity)
  $= E[E(\varepsilon X|X)] = 0$

Regression & LS

Consider AR: $y_t = \rho\ y_{t-1} + \varepsilon_t, \quad \varepsilon_t\ i.i.d.\ N(0, \sigma^2)$

Notice that $E(y_t, \varepsilon_t) = E[(\rho\ y_{t-1} + \varepsilon_t)\varepsilon_t] =$
$\rho\ E(y_{t-1}\varepsilon_t) + E(\varepsilon_t^2) = \rho\ E(y_{t-1}\varepsilon_t) + \sigma^2 = \sigma^2 \neq 0$
$(E(y_{t-1}\varepsilon_t) = 0)$
i.e. Strict Exogeneity doesn't hold for TS data.

H_0 : $\beta = \overline{\beta}$ or $R\beta = r$ v.s. H_1 : $\beta_k \neq \overline{\beta}_k$ at least for some k.

$$F = \frac{(SSR_0 - SSR_1)/k \rightarrow \#r}{SSR_1\ /\ (n-k)} \sim F_{k,n-k} \sim \chi^2_k \text{ as } n \rightarrow \infty$$
$$= \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$