Lecture 24 - 10/19

Friday, October 21, 2022 9:05 AM

## Review:

① Guassian distribution

- $X$ has guassian (normal) distribution, $X \sim N(\mu, \sigma^2)$ if

$$f_X(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right), \quad u \in \mathbb{R}$$

$$E[X] = \mu, \quad Var(X) = \sigma^2$$

- Standardized version of $X$, $\tilde{X} = \frac{X-\mu}{\sigma}$ is distributed as $N(0,1)$.

  - $F_{\tilde{X}}(a) = P(\tilde{X} \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{u^2}{2}\right) du = \Phi(a)$

  - $F^c_{\tilde{X}}(a) = P(\tilde{X} > a) = \int_{a}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = 1 - \Phi(a) = Q(a)$

  - Notice that $Q(a) = \Phi(-a)$ because of symmetry!

② Guassian approximation with error correction

Let $S \sim Bi(n,p)$ denote a random variable with Binomial distribution. Let $X \sim N(\mu, \sigma^2)$ denote a Guassian random variable with same mean and variance as $S$, i.e., $\mu = np, \sigma^2 = np(1-p)$

$$\left.\begin{aligned} P(S \leq k) &\approx P(X \leq k + 0.5) \\ P(S > k) &\approx P(X \geq k + 0.5) \end{aligned}\right\} \text{ sum of these two should be one}$$

$$P(S \geq k) \approx P(X \geq k - 0.5)$$

③ ML parameter estimation for continuous-type random variables

(i) pdf of random variable $X$ belongs to a family of parametrized distributions: $f_\theta$

(ii) We are given an observation $X = u$.

(iii) We want to find the parameter that maximizes the pdf at $u$, i.e., $\hat{\theta}_{ML} = \underset{\theta}{\arg\max}\, f_\theta(u)$

## Today:

① The distribution of a function of a random variable

---

① The distribution of a function of a random variable

Suppose that $X$ is a continuous-type random variable and $Y = g(X)$ for some function $g$.

Our goal is to find distribution of $Y$.

Step 1: Find whether $Y$ is continuous-type or discrete-type and find its support.

**Step 1**: Find whether $Y$ is continuous-type or discrete-type and find its support.

To do this, use the following steps:

(i) sketch pdf of $X$ and function $g$ on the same plot.

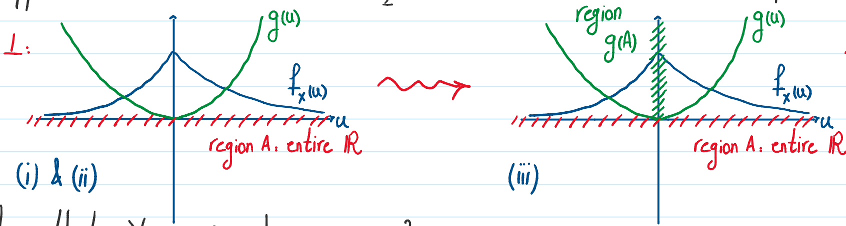
(ii) identify support of $X$, i.e., region over $x$-axis for which $f_X$ is positive. Let's call this region $A$.

(iii) identify $g(A)$ on the $y$-axis. This is support of $Y$. If $g(A)$ is itself a continuous region, $Y$ is continuous-type. If $g(A)$ is union of points, it is discrete type.

note: Sometimes, step 1 is easy. You can identify support of $Y$ without using above steps.

example: Suppose that $Y=X^2$, and $f_X(u)=\frac{1}{2}e^{-|u|}$ for $u\in\mathbb{R}$. We want to find $f_Y(u)$.

Step 1:




$\rightsquigarrow$ $Y$ is continuous type since $g(A)$ is a continuous region. Support of $Y$ is $g(A)$, i.e., $u\geq 0$

Notice that $Y=g(X)$ where $g(u)=u^2$

→ If $Y$ is continuous type:

**Step 2**: Find CDF of $Y$ using $F_Y(c)=P(Y\leq c)=P(g(X)\leq c)=\int_{u:\,g(u)\leq c} f_X(u)\,du$.

note: If it is possible, write down $F_Y$ in terms of $F_X$. This way you do not need to do integration

example: Step 2. So far, we have that support of $Y$ is positive values. For any $c>0$

$$F_Y(c)=P(X^2\leq c)=P(-\sqrt{c}\leq X\leq\sqrt{c})=P(X\leq\sqrt{c})-P(X\leq-\sqrt{c})=F_X(\sqrt{c})-F_X(-\sqrt{c})$$

**Step 3**: Find pdf of $Y$ by differentiating $F_Y$, i.e., $f_Y(u)=\frac{dF_Y(u)}{du}$

example: $\frac{dF_Y(u)}{du}=\frac{d}{du}\left(F_X(\sqrt{u})-F_X(-\sqrt{u})\right)=\frac{1}{2\sqrt{u}}f_X(\sqrt{u})+\frac{1}{2\sqrt{u}}f_X(-\sqrt{u})=\frac{1}{2\sqrt{u}}e^{-\sqrt{u}}$ for $u\geq 0$

→ If $Y$ is discrete-type:

**Step 2**: Find pmf of $Y$ using $P(Y=k)=P(g(X)=k)=\int_{u:\,g(u)=k} f_X(u)\,du$

---

Some remarks:

- After finding distribution of $Y$, you can use it to get statistical quantities related to $Y$ using $f_Y$ or LOTUS.

example: for the above example:

$$E[Y]=E[g(X)]=E[X^2]=\frac{1}{2}\int_{-\infty}^{+\infty}u^2e^{-|u|}=\int_0^\infty u^2e^{-u}\,du=2!$$ $\rightsquigarrow$ integration by part + LOTUS

$$E[Y]=\int_{-\infty}^{+\infty}u\,f_Y(u)\,du=\int_0^\infty u\cdot\frac{1}{2\sqrt{u}}e^{-\sqrt{u}}\,du=\int_0^\infty\frac{\sqrt{u}}{2}e^{-\sqrt{u}}\,du=\int_0^\infty v^2e^{-v}\,dv=2!$$ $\rightsquigarrow$ change of variable $v=\sqrt{u}$

- Although we assumed $X$ is continuous-type the same can be used if $X$ is a discrete random variable

• Although we assumed $X$ is continuous-type the same can be used if $X$ is a discrete random variable

**Example 3.8.11** Suppose $X$ is a continuous-type random variable with CDF $F_X$. Let $Y$ be the result of applying $F_X$ to $X$, that is, $Y = F_X(X)$. Find the distribution of $Y$.

Solution:

Step 1. Notice that $Y = g(X) = F_X(X)$. Hence $g = F_X$. We know range of $F_X$ is $[0,1]$ and $X$ is continuous type.
Hence, support of $Y$ is $[0,1]$

Step 2: $P(Y \leq v) = P(F_X(X) \leq v)$, $v \in [0,1]$

Since $X$ is continuous type, $F_X$ is increasing. Moreover for any $v \in (0,1)$ there exists a unique $c_v$ such that $F_X(c_v) = v$. Hence, for any $v \in (0,1)$ we have $P(F_X(X) \leq v) = P(X \leq c_v) = F_X(c_v) = v$.

Notice that $P(Y \leq 0) = 0$ and $P(Y \leq 1) = 1$

Step 3: $f_Y(u) = \frac{dF_Y(u)}{du} = 1$ for $u \in [0,1]$

Hence $Y$ is a uniformly distributed random variable over $[0,1]$.

Remark:

• Suppose that $Y \sim Unif([0,1])$ & $F$ is cdf of a continuous type random variable. The above example suggests $X = F^{-1}(Y)$ is a continuous type random variable with cdf $F_X = F$
Hence, if we have a method for generating $Y \sim Unif([0,1])$ random variable, we can generate any continuous-type random variable with cdf $F$, by pass $Y$ through $F^{-1}$