14.381 RECITATION 09/24/2004

HANDOUT #1

TA: José Tessada
(tessada@mit.edu)

1. TRANSFORMATION OF A R.V.

Thm. 2.1.3 (p. 51) Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x : f_X(x) > 0\}$$
$$\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}.$$

(a) If $g(\cdot)$ is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$

(b) If $g(\cdot)$ is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

Example 2.1.4

$$X \sim f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{Uniform } (0,1) \text{ dist.}$$

$\Rightarrow F_X(x) = x, \quad 0 < x < 1$

define $Y = g(X) = -\log X$

notice that $g(x)$ is decreasing for $x \in (0,1)$.
The range of Y is between 0 and $\infty \Rightarrow \mathcal{Y} = (0, \infty)$

$g^{-1}(y) = e^{-y}$

Then for $y \in \mathcal{Y}$ $(y > 0)$

$$F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}$$

and $F_Y(y) = 0, \quad y \leq 0$.

Be careful: we need monotonicity on the range of X!

This is very useful when the functions are monotone on the whole support of X.

The theorem as stated cannot be directly applied to a function that is not monotone on the whole support. Theorem 2.1.5 follows from Thm. 2.1.3, it is the equivalent for pdf.

When $g(\cdot)$ is not monotone on the whole support of X but it is monotone over certain intervals we can generalize Thm. 2.1.5.
(As the book says: in the case of a $g(\cdot)$ that <u>is not</u> monotone over certain intervals, <u>you</u> are in DEEP trouble.)

The following thm. is an extension of 2.1.5 and will be very useful in many cases:

<u>Theorem 2.1.8</u> Let X have pdf $f_X(x)$, let $Y = g(X)$, and define the sample space $\mathcal{X} = \{x : f_X(x) > 0\}$. Suppose there exists a partition, $A_0, A_1, \dots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \dots, g_k(x)$, defined on $A_1, \dots, A_k$, respectively, satisfying

i. $g(x) = g_i(x)$, for $x \in A_i$,
ii. $g_i(x)$ is monotone on A_i,
iii. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $i = 1, \dots, k$, and
iv. $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$, for each $i = 1, \dots, k$.

Then,

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise} \end{cases}$$

Note that $g_i(x)$ is a one-to-one transformation from A_i onto $\mathcal{Y}$. $g_i^{-1}(y)$ is one-to-one too (from $\mathcal{Y}$ onto A_i).

Example 2.1.9 Normal - χ^2 relationship

Let $X \sim N(0,1)$ $\qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad -\infty < x < \infty$

Consider $Y = X^2$.
Then, $\mathcal{Y} = (0, \infty)$.
If we take

$A_0 = \{0\}$
$A_1 = (-\infty, 0) \qquad g_1(x) = x^2 \qquad g_1^{-1}(y) = -\sqrt{y}$
$A_2 = (0, \infty) \qquad g_2(x) = x^2 \qquad g_2^{-1}(y) = \sqrt{y}$

we can apply thm. 2.1.8 to obtain

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\{-(-\sqrt{y})^2/2\} \left| -\frac{1}{2\sqrt{y}} \right|$$

$$+ \frac{1}{\sqrt{2\pi}} \exp\{-(\sqrt{y})^2/2\} \left| \frac{1}{2\sqrt{y}} \right|$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, \qquad 0 < y < \infty$$

Y is said to have a χ^2 distribution with one degree of freedom (χ^2_p or $\chi^2(p)$)! $p=1$ This distribution will become a very "close friend" of us later in the course.

Note: Theorem 2.1.10 $\longrightarrow$ very useful for "random sample" generation. (univariate distribution - bivariate distributions require more elaboration).

2. MOMENT GENERATING FUNCTION

Exercise 2.32 Let $M_X(t)$ be the mgf of X, and define $S(t) = \log M_X(t)$. Show that

$$\frac{d}{dt} S(t)\Big|_{t=0} = E(X) \quad \text{and} \quad \frac{d^2}{dt^2} S(t)\Big|_{t=0} = Var(X).$$

Solution:

$$\frac{d}{dt} S(t)\Big|_{t=0} = \frac{d}{dt}\left(\log M_x(t)\right)\Big|_{t=0} = \frac{\frac{d}{dt} M_x(t)}{M_x(t)}\Bigg|_{t=0}$$

$$= \frac{E(X)}{1} = E(X)$$

where we made use of the fact that $M_x(0) = 1 \left(= E(e^0)\right)$

$$\frac{d^2}{dt^2} S(t)\Big|_{t=0} = \frac{d}{dt}\left(\frac{M_x'(t)}{M_x(t)}\right)\Bigg|_{t=0} = \frac{M_x(t) M_x''(t) - [M_x'(t)]^2}{[M_x(t)]^2}\Bigg|_{t=0}$$

$$= \frac{1 \cdot E(X^2) - E(X)^2}{1} = Var(X)$$

Exercise 2.33 (b)

$$P(X=x) = p(1-p)^x, \qquad M_X(t) = \frac{p}{1-(1-p)e^t} \qquad x = 0, 1, \ldots \quad 0<p<1$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x = p \sum_{x=0}^{\infty} ((1-p)e^t)^x = p \frac{1}{1-(1-p)e^t}$$

$$= \frac{p}{1-(1-p)e^t}, \qquad t < -\log(1-p)$$

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

$$E(X^2) = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0}$$
$$= \frac{(1-(1-p)e^t)^2 (p(1-p)e^t) + p(1-p)e^t\, 2(1-(1-p)e^t)(1-p)e^t}{(1-(1-p)e^t)^4}\Bigg|_{t=0}$$
$$= \frac{p^3(1-p) + 2p^2(1-p)^2}{p^4} = \frac{p(1-p) + 2(1-p)^2}{p^2}$$

$$\mathrm{Var}(X) = \frac{p(1-p) + 2(1-p)^2}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2} \quad //$$

3. DIFFERENTIATING UNDER AN INTEGRAL SIGN

Theorem 2.4.1 (Leibnitz's Rule) If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable w.r.t. θ, then

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x,\theta)\, dx = f(b(\theta),\theta) \frac{d}{d\theta} b(\theta) - f(a(\theta),\theta) \frac{d}{d\theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\, dx \qquad \blacksquare$$

Section 2.4, C&B (p. 68–75) covers this topic and also the interchange of summation and differentiation.

A general result is that if we are integrating over a finite range or we have a finite sum, then differentiation is not a problem.

This can be easily seen using theorem 2.4.1 if we take $a(\theta)$ and $b(\theta)$ to be constants. In this case we obtain the following particular case of Leibnitz's Rule:

$$\frac{d}{d\theta} \int_a^b f(x,\theta)\, dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\, dx \qquad (1)$$

One student pointed out the fact that in ~~[illegible]~~ one of the formula presented in the lecture there were ordinary derivatives and partial derivatives. The same thing is observed in (1)! Before integrating $f(x,\theta)$ is a function of both x and θ, so the derivative is a partial derivative; after integrating (LHS) we have that $h(\theta) = \int_a^b f(x,\theta)\,dx$ is a function of θ only $\Rightarrow$ we take ordinary derivatives.

Example: think of a random variable X with a $N(\mu, \sigma^2)$ - I am not going to proof any result for interchanging derivatives and integrals! - we know that if $X \sim N(\mu, \sigma^2)$ then the pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-(x-\mu)^2/2\sigma^2\right\}$$

Note that in fact we can write $f_X(x, \theta)$ where $\theta = (\mu, \sigma^2)$.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x,\theta)\,dx = \mu$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x,\theta)\,dx = \mu^2 + \sigma^2$$

$$E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x,\theta)\,dx = \sigma^2$$

As you can clearly see all the expectations are functions of θ, but x doesn't appear as an argument. In fact, (*) as we integrated over x, we "eliminated" the variable

Back to the differentiation and integration discussion, theorems 2.4.2 and 2.4.3 in the book can be interpreted as () smoothness conditions (although some uniformity is required too).

(*) This will show up again in the study of () multiple random variables.