Stat 2911 Lecture Notes

Class 32, 2017

Uri Keich

Prediction, MGFs, moments and derivatives of the MGF (Rice 4.4, 4.5)

Prediction (of unobserved, not necessarily future events)

Suppose X and Y are jointly distributed RVs.

The prediction problem is to minimize the error, find:

$$\underset{g:\mathbb{R}\to\mathbb{R}}{\operatorname{argmin}} E[Y-g(X)]^2.$$

$$E\{[Y-g(X)]^2\} = E\{E([Y-g(X)]^2 \mid X)\}$$

$$\underbrace{E([Y-g(X)]^2 \mid X)}_{W} \geq \underbrace{V(Y \mid X)}_{U} \quad \text{with "=" iff } g(X) = E(Y \mid X)$$

Claim. If $W, U \in L^1$ and $W \geq U$, then

$$E(W) \geq E(U) \quad \text{with "=" iff } P(W=U)=1.$$

$$\Rightarrow \quad E\{[Y-g(X)]^2\} = E\{E([Y-g(X)]^2 \mid X)\} = E(W)$$

$$\geq E(U) = E[V(Y \mid X)]$$

with "=" iff $W=U$ a.s. $\Longleftrightarrow$ $g(X) = E(Y \mid X)$ a.s.

Cor. The optimal MSE predictor of Y given X is $E(Y \mid X)$ and the MSE in this case is $E[V(Y \mid X)]$.

Note: if Y is ind. of X then $E(Y \mid X) =$ (same as when we don't know X), and $MSE = V(Y)$ (why?)

Example. Z and W are standard bivariate normal with parameter / correlation coefficient $\rho \in (-1, 1)$.

Recall: $Z \mid W = w \sim N(\rho w, 1-\rho^2)$

$$\Rightarrow E(Z \mid W = w) =$$

$$\Rightarrow E(Z \mid W) =$$

$\Rightarrow \rho \cdot W$ minimizes $E[(Z - g(W))^2]$ among all functions g.

Notably, the best MSE predictor of Z given W is linear in W. Again, this is a property of the multivariate normal dist. and is not the case generally.

$$MSE = E[(Z - \rho W)^2]$$
$$= E(Z^2) - 2\rho E(ZW) + \rho^2 E(W^2)$$
$$= 1 - \rho^2$$

Compare this with $V(Z \mid W) =$ (invariant of W, again, a property of the normal)

$$\Rightarrow \quad E[V(Z \mid W)] =$$

If $\rho = 0$, then $E(Z \mid W) =$

$$= E(Z) \quad \text{(Z, W are ind.)}$$

$$MSE = \quad = V(Z)$$

As $|\rho| \uparrow 1$, $MSE \downarrow 0$.

Claim. If $W, U \in L^1$ and $W \geq U$, then

$$E(W) \geq E(U) \text{ with "=" iff } P(W=U)=1.$$

Proof.

Let $X = W-U$, then $X \geq 0$.

(intuitively) $\Rightarrow$ $E(X) \geq 0$ (tricky if not jointly discrete/cont.)

Fact: if $X \geq 0$ $E(X) \overset{(*)}{=} \int_0^\infty (1-F_X(x))\,dx$ (proved for X disc./cont.)

$$\Rightarrow E(W)-E(U) = E(X) \geq 0.$$

If $W=U$ a.s. then $X = W-U = 0$ a.s. so $\forall x \geq 0$

$$1 \geq F_X(x) = P(X \leq x) \geq P(X=0) = 1$$

$$(*) \Rightarrow \quad E(X) = 0$$

$$\Rightarrow \quad E(W) = E(U).$$

Conversely, if $E(W) = E(U)$ then

$$0 = E(X) = \int_0^\infty (1-F_X(x))\,dx \Rightarrow 1-F_X(x) = 0 \quad \forall x > 0$$

Since $X \geq 0$, $F_X(x) = 0 \ \forall x < 0$ so $F_X(x) = \begin{cases} 0 & x<0 \\ 1 & x \geq 0 \end{cases}$

(F is right cont.)

$\Rightarrow X = 0$ a.s. $\Rightarrow W = U$ a.s.

<u>Moment Generating Function</u> (MGF)

Let X be a RV, then $Y = e^{tX} \geq 0$, so EY is well-defined, albeit it can be $+\infty$.

The MGF of the RV X is defined as

$$M(t) = Ee^{tX} \quad \forall t \text{ for which the RHS} < \infty$$

$$M(0) = Ee^{0 \cdot X} = ,$$

so the MGF is always finite for $t=0$.

If X is a discrete RV with pmf $\sum_i' p_X(x_i) = 1$ then

$$M(t) = \sum_i' e^{tx_i} p_X(x_i) \quad (\text{where} < \infty).$$

If X is a cont. RV with density f_X then

$$M(t) = \int_{\mathbb{R}} e^{tx} f_X(x)\,dx \quad (\text{where} < \infty).$$

<u>Examples</u>. i) $X \sim Pois(\lambda)$ where $\lambda > 0$:

$$M(t) = \sum_{k=0}^{\infty}{}' e^{tk} e^{-\lambda} \lambda^k / k!$$

$$= e^{-\lambda} \sum_{k=0}^{\infty}{}' (\underbrace{\lambda e^t}_{x})^k / k!$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \quad \forall t \in \mathbb{R}$$

(ii) $X \sim N(0,1)$

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2tx + t^2) + t^2/2}\, dx$$

$$= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2}\, dx$$

$$= e^{t^2/2} \qquad \forall t \in \mathbb{R}$$

(iii) $X \sim \Gamma(\alpha, \lambda) \qquad \alpha, \lambda > 0$

$$M(t) = \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\lambda - t)x}\, dx \quad \begin{cases} & \text{if } t \geq \lambda \\ & \text{if } t < \lambda \end{cases}$$

For $t < \lambda$ we have

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha \int_0^{\infty} \frac{(\overbrace{\lambda - t}^{\nu})^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\overbrace{\lambda - t}^{\nu})x}\, dx \qquad \nu > 0$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^\alpha \qquad \forall t < \lambda.$$

MGFs are useful for two reasons.

Theorem M1. If $\exists \delta > 0$ s.t. $M(t) < \infty$ $\forall t \in (-\delta, \delta)$ then $\forall n \in \mathbb{N}$ $M^{(n)}(0) = E(X^n)$ (and both exist)

Intuitively, it makes sense:

$$M'(t) = \frac{d}{dt} \sum_i e^{tx_i} p_X(x_i)$$

$$\overset{??}{=} \sum_i \frac{d}{dt} e^{tx_i} p_X(x_i)$$

$$= \sum_i x_i e^{tx_i} p_X(x_i)$$

$$\Rightarrow M'(0) = \sum_i x_i p_X(x_i) =$$

Similarly,

$$M^{(k)}(t) = \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$

$$\overset{??}{=} \int_{-\infty}^{\infty} \frac{d^k}{dt^k} e^{tx} f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^k e^{tx} f_X(x)\, dx$$

$$\Rightarrow M^{(k)}(0) = \int_{-\infty}^{\infty} x^k f_X(x)\, dx =$$

Examples

i) $X \sim \mathrm{Pois}(\lambda)$ $\qquad M(t) = e^{\lambda(e^t - 1)} \quad \forall t \in \mathbb{R}$

$$M'(t) = M(t)\,\lambda e^t$$

$$\Rightarrow \quad M'(0) = M(0)\cdot\lambda = \qquad = E(X)$$

$$M''(t) = M(t)(\lambda e^t)^2 + M(t)\lambda e^t$$

$$\Rightarrow \quad M''(0) = \qquad = E(X^2)$$

(ii) $X \sim N(0,1)$ $\qquad M(t) = e^{t^2/2} \quad \forall t \in \mathbb{R}$

$$M'(t) = M(t)\cdot t$$

$$\Rightarrow \quad M'(0) = \qquad = E(X)$$

$$M''(t) = M(t)\cdot t^2 + M(t)$$

$$\Rightarrow \quad M''(0) = \qquad = E(X^2)$$

(iii) $X \sim \Gamma(\alpha,\lambda)$ $\qquad M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} = \lambda^{\alpha}(\lambda - t)^{-\alpha} \quad \forall t < \lambda$

$$M'(t) = \lambda^{\alpha}(\lambda - t)^{-\alpha-1}(-\alpha)(-1)$$

$$= \alpha\,\lambda^{\alpha}(\lambda - t)^{-(\alpha+1)} \qquad t < \lambda$$

$$\Rightarrow \quad M'(0) = \alpha\,\lambda^{\alpha}\,\lambda^{-(\alpha+1)} = \frac{\alpha}{\lambda} = E(X)$$

$$M''(t) = \alpha(\alpha+1)\lambda^{\alpha}(\lambda - t)^{-(\alpha+2)} \qquad t < \lambda$$

$$\Rightarrow \quad M''(0) = \frac{\alpha(\alpha+1)}{\lambda^2} = E(X^2)$$

<u>Theorem M2</u>. Let F and G be CDFs and suppose $\exists \delta > 0$ s.t. $\forall t \in (-\delta, \delta)$ $M_F(t) = M_G(t)$ $(< \infty)$. Then $F \equiv G$.

It follows that all the moments of F and G exist and are equal! The converse is <u>false</u>: all the moments of F and G can exist, and be equal, yet $F \not\equiv G$.