STAT 450 Lecture 10

Reading for Today's Lecture: ?
Goals of Today's Lecture:

- Discuss Monte Carlo techniques
- Introduce statistical inference

Expectation, moments

We give two definitions of expected values:
Def'n If X has density f then

$$E(g(X)) = \int g(x)f(x)\,dx\,.$$

Def'n: If X has discrete density f then

$$E(g(X)) = \sum_x g(x)f(x)\,.$$

Now if $Y = g(X)$ for smooth g and X has density f_X then

$$\begin{aligned} E(Y) &= \int y f_Y(y) \\ &= \int g(x) f_Y(g(x)) g'(x)\,dx \\ &= \int g(x) f_X(x)\,dx \\ &= E(g(X)) \end{aligned}$$

by the change of variables formula for integration. This is good because otherwise we might have two different values for $E(e^X)$.

In general, there are random variables which are neither absolutely continuous nor discrete. Look at my STAT 801 web pages to see how E is defined in general.

Facts: E is a linear, monotone, positive operator:

1. **Linear**: $E(aX + bY) = aE(X) + bE(Y)$ provided X and Y are integrable.
2. **Positive**: $P(X \geq 0) = 1$ implies $E(X) \geq 0$.

3. **Monotone**: $P(X \geq Y) = 1$ and X, Y integrable implies $E(X) \geq E(Y)$.

Major technical theorems:

Monotone Convergence: If $0 \leq X_1 \leq X_2 \leq \cdots$ and $X = \lim X_n$ (which has to exist) then

$$E(X) = \lim_{n \to \infty} E(X_n)$$

Dominated Convergence: If $|X_n| \leq Y_n$ and there is a random variable X such that $X_n \to X$ (technical details of this convergence later in the course) and a random variable Y such that $Y_n \to Y$ with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \to E(X)$$

This is often used with all Y_n the same random variable Y.

Fatou's Lemma: If $X_n \geq 0$ then

$$E(\limsup X_n) \leq \limsup E(X_n)$$

Theorem: With this definition of E if X has density $f(x)$ (even in R^p say) and $Y = g(X)$ then

$$E(Y) = \int g(x) f(x) dx \, .$$

(This could be a multiple integral.) If X has pmf f then

$$E(Y) = \sum_x g(x) f(x) \, .$$

This works for instance even if X has a density but Y doesn't.

Def'n: The r^{th} moment (about the origin) of a real random variable X is $\mu'_r = E(X^r)$ (provided it exists). We generally use μ for $E(X)$. The r^{th} central moment is

$$\mu_r = E[(X - \mu)^r]$$

We call $\sigma^2 = \mu_2$ the variance.

Def'n: For an R^p valued random vector X we define $\mu_X = E(X)$ to be the vector whose i^{th} entry is $E(X_i)$ (provided all entries exist).

Def'n: The $(p \times p)$ variance covariance matrix of X is

$$Var(X) = E\left[(X - \mu)(X - \mu)^t\right]$$

which exists provided each component X_i has a finite second moment.

Moments and probabilities of rare events are closely connected as will be seen in a number of important probability theorems. Here is one version of Markov's inequality (one case is Chebyshev's inequality):

$$\begin{aligned} P(|X-\mu| \geq t) &= E[1(|X-\mu| \geq t)] \\ &\leq E\left[\frac{|X-\mu|^r}{t^r} 1(|X-\mu| \geq t)\right] \\ &\leq \frac{E[|X-\mu|^r]}{t^r} \end{aligned}$$

The intuition is that if moments are small then large deviations from average are unlikely.

Example moments: If Z is standard normal then

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \frac{-e^{-z^2/2}}{\sqrt{2\pi}}\Bigg|_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

and (integrating by parts)

$$\begin{aligned} E(Z^r) &= \int_{-\infty}^{\infty} z^r e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \frac{-z^{r-1} e^{-z^2/2}}{\sqrt{2\pi}}\Bigg|_{-\infty}^{\infty} + (r-1)\int_{-\infty}^{\infty} z^{r-2} e^{-z^2/2} dz/\sqrt{2\pi} \end{aligned}$$

so that

$$\mu_r = (r-1)\mu_{r-2}$$

for $r \geq 2$. Remembering that $\mu_1 = 0$ and

$$\mu_0 = \int_{-\infty}^{\infty} z^0 e^{-z^2/2} dz/\sqrt{2\pi} = 1$$

we find that

$$\mu_r = \begin{cases} 0 & r \text{ odd} \\ (r-1)(r-3)\cdots 1 & r \text{ even} \end{cases}$$

If now $X \sim N(\mu, \sigma^2)$, that is, $X \sim \sigma Z + \mu$, then $E(X) = \sigma E(Z) + \mu = \mu$ and

$$\mu_r(X) = E[(X - \mu)^r] = \sigma^r E(Z^r)$$

In particular, we see that our choice of notation $N(\mu, \sigma^2)$ for the distribution of $\sigma Z + \mu$ is justified; σ is indeed the variance.

If $X \sim MVN(\mu, \Sigma)$ then we have $X = AZ + \mu$ with Z standard multivariate normal. Hence

$$\begin{aligned} \mathrm{E}(X) &= A\mathrm{E}(Z) + \mu \\ &= A \cdot 0 + \mu \\ &= 0 \end{aligned}$$

Moreover

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}\left[(X - \mu)(X - \mu)^T\right] \\ &= \mathrm{E}\left[(AZ)(AZ)^T\right] \\ &= A\mathrm{E}\left[ZZ^T\right] A^T \end{aligned}$$

To compute $\mathrm{E}\left[ZZ^T\right]$ look at entry ij in ZZ^T which is $Z_i Z_j$. Then

$$\begin{aligned} \mathrm{E}(Z_i Z_j) &= \begin{cases} \mathrm{E}(Z_i^2) & i = j \\ \mathrm{E}(Z_i)\mathrm{E}(Z_j) & \text{1} \neq j \end{cases} \\ &= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \end{aligned}$$

SO: $\mathrm{E}(ZZ^T) = I_{n \times n}$ and

$$\mathrm{Var}(X) = AA^T = \Sigma\,.$$

Moments and independence

Theorem: If $X_1, \ldots, X_p$ are independent and each X_i is integrable then $X = X_1 \cdots X_p$ is integrable and

$$E(X_1 \cdots X_p) = E(X_1) \cdots E(X_p)$$