Statistical Properties

Richard Yi Da Xu

School of Computing & Communication, UTS

December 6, 2015

Change in variables

Let $Y = g(X)$:

$$
\begin{aligned}
F_Y(y) &= P(g(X) \leq y) \\
&= P(X \leq g^{-1}(y)) \\
&= F_X(g^{-1}(y))
\end{aligned}
$$

$$
\begin{aligned}
F_Y(y) = \frac{\partial F_Y(y)}{\partial y} &= \frac{F_X(g^{-1}(y))}{\partial x} \times \frac{\partial x}{\partial y} \\
&= f_x(g^{-1}(y)) \frac{\partial g^{-1}(y)}{\partial y} \\
&= f_x(g^{-1}(y)) \left| \frac{\partial g^{-1}(y)}{\partial y} \right|
\end{aligned}
$$

Some thoughts on Floor function

- We let $X \sim \text{Exp}(1)$, i.e., $X \sim e^{-x}$.
- $Y = \lfloor X \rfloor$, and
- $Z = X - \lfloor X \rfloor$

$$
\begin{aligned}
\Pr(Y = \lfloor x \rfloor) &= P(X < (x+1)) - P(X < x) \\
&= \int_{t=0}^{x+1} e^{-t} \mathrm{d}t - \int_{t=0}^{x} e^{-t} \mathrm{d}t \\
&= e^{-x} - e^{-(x+1)} = e^{-x}(1 - e^{-1})
\end{aligned}
$$

Let $0 \leq z \leq 1$:

$$
\begin{aligned}
\Pr(Z < z) &= \int_{t=0}^{z} \sum_{i=0}^{\infty} p(t + i = z + i) \mathrm{d}t \\
&= \int_{t=0}^{z} \sum_{i=0}^{\infty} e^{-(t+i)} \mathrm{d}t = \sum_{i=0}^{\infty} \int_{t=0}^{z} e^{-(t+i)} \mathrm{d}t \\
&= \sum_{i=0}^{\infty} \left[-e^{-(t+i)} \right]_{t=0}^{z} = \sum_{i=0}^{\infty} e^{-i} - e^{-(z+i)} \\
&= \sum_{i=0}^{\infty} e^{-i}(1 - e^{-z}) = (1 - e^{-z}) \sum_{i=0}^{\infty} e^{-i} = (1 - e^{-z}) \frac{1 - (\exp^{-\infty})}{1 - e^{-1}} \\
&= \frac{1 - e^{-z}}{1 - e^{-1}} \qquad \text{A valid CDF, as } \Pr(Z \leq 0) = 0 \text{ and } \Pr(Z \leq 1) = 1
\end{aligned}
$$

- Therefore, $P(Z < z)$ is independant of $P(Y = \lfloor x \rfloor)$, as it does NOT contain x terms.

Markovs inequality
Let X be a nonnegative random variable. Then, for any $b \in \mathbb{R}^+$:

$$\Pr(X \geq b) \leq \frac{\mathbb{E}[X]}{b}$$

Why?

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty xp(x)\mathrm{d}x = \int_0^b xp(x)\mathrm{d}x + \int_b^\infty xp(x)\mathrm{d}x \\
\implies \mathbb{E}[X] &\geq \int_b^\infty xp(x)\mathrm{d}x \\
&\geq \int_b^\infty bp(x)\mathrm{d}x \\
&= b\int_b^\infty p(x)\mathrm{d}x \\
&= b\Pr(X \geq b)
\end{aligned}$$

how is this useful? provides an **upper bound** of probability that a nonnegative random variable is greater than an arbitrary positive constant by relating a probability to an expectation.

Useful inequalities: Chebyshev's inequality

Let X be a nonnegative random variable. Then, for any $b \in \mathbb{R}^+$:

$$\Pr(X \geq b) \leq \frac{\mathbb{E}[X]}{b}$$

substitute $X \rightarrow (X - \mu)^2$ and $b \rightarrow k^2$:

$$\begin{aligned}
&\implies \Pr\left((X - \mu)^2 \geq k^2\right) \leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2} \\
&\implies \Pr(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}
\end{aligned}$$

substitute $X \rightarrow (X - \mu)^2$ and $b \rightarrow \sigma^2 k^2$:

$$\begin{aligned}
&\implies \Pr\left((X - \mu)^2 \geq \sigma^2 k^2\right) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\sigma^2 k^2} = \frac{1}{k^2} \\
&\implies \Pr(|X - \mu| \geq \sigma k) \leq \frac{1}{k^2}
\end{aligned}$$

Chebyshev's inequality applications (1)

Provides bounds of random variables from any distributions when their means and variances are known. Each k tells us one bound, for example, when $k = 2$:

$$
\begin{aligned}
&\Pr(|X - \mu| \geq 2\sigma) \leq \frac{1}{4} \implies \Pr(\mu - X \geq 2\sigma, \mu + X \geq 2\sigma) \leq \frac{1}{4} \\
\implies &\Pr(X \leq \mu - 2\sigma, X \geq \mu + 2\sigma) \leq \frac{1}{4} \\
\implies &\Pr(\mu - 2\sigma \leq X \geq \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4}
\end{aligned}
$$

For Guassian distribution, $\Pr(\mu - 2\sigma \leq X \geq \mu + 2\sigma) \approx 0.995$

Chebyshev's inequality applications (2)

Let $X_n \in \text{Gamma}(n, \frac{1}{n})$, therefore:

$$\mathbb{E}[X_n] = n \times \frac{1}{n} = 1 \qquad \mathbb{VAR}[X_n] = n \times \left(\frac{1}{n}\right)^2 = \frac{1}{n}$$

Therefore,

$$\begin{aligned} &\Pr\left(|X - \mu| > k\right) \leq \frac{\sigma^2}{k^2} \\ \implies &\Pr\left(|X_n - 1| > \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} = \frac{1}{n\epsilon^2} \to 0 \qquad \text{as } n \to \infty \end{aligned}$$

Definition X_n **converges in probability** to the random variable X i.e., $X_n \xrightarrow{P} X$:

$$\Pr(|X_n - X| > \epsilon) \to 0 \qquad \text{as } n \to \infty$$

law of large numbers

- Let $X1, X2, \dots X_n$ be a sequence of i.i.d. random variables with mean μ and finite variance σ^2
- Let $S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$.
  $\mathbb{VAR}(aX + bY) = a^2 \mathbb{VAR}(X) + b^2 \mathbb{VAR}(Y)$ if X and Y are independant

$$\mathbb{E}[S_n] = \frac{\mathbb{E}[X_1]}{n} + \cdots + \frac{\mathbb{E}[X_n]}{n} = \mu$$
$$\mathbb{VAR}[S_n] = \frac{\mathbb{VAR}[X_1]}{n^2} + \cdots + \frac{\mathbb{VAR}[X_n]}{n^2} = \frac{\sigma^2}{n}$$

Therefore,

$$\Pr(|X - \mu| > \epsilon) \leq \frac{\mathbb{VAR}[S_n]}{\epsilon^2} \implies \Pr\Big(|S_n - \mu| > \epsilon\Big) \leq \frac{\sigma^2}{n\epsilon^2}$$
$$\implies \Pr\Big(|S_n - \mu| > \epsilon\Big) \to 0 \qquad \text{as } n \to \infty$$

- The law of large numbers states that $S_n \xrightarrow{P} \mu$

$$\Pr(|S_n - \mu| > \epsilon) \to 0 \qquad \text{as } n \to \infty$$

Uniqueness in almost surely Convergence (1)

- X_n converges almost surely (a.s.) to the random variable X as $n \to \infty$ iff:

$$P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1 \qquad \text{as } n \to \infty$$

- Let $X_1, X_2, \dots X_n$ be a sequence of random variables. If X_n converges almost surely, then the limiting random variable (distribution) X is unique.
- Suppose **not** unique: $X_n \xrightarrow{\text{a.s}} X$ and $X_n \xrightarrow{\text{a.s}} Y$ as $n \to \infty$.

$$\begin{aligned}
\Phi_X &= \{\omega : X_n(\omega) \not\to X(\omega) \text{ as } n \to \infty\} \text{ and} \\
\Phi_Y &= \{\omega : Y_n(\omega) \not\to X(\omega) \text{ as } n \to \infty\}
\end{aligned}$$

- First we proved that $\omega \notin (\Phi_X \cup \Phi_Y) \implies X(\omega) = Y(\omega)$
- Since $\omega \in (\Phi_X^c \cap \Phi_Y^c) \equiv \omega \notin (\Phi_X \cup \Phi_Y) \equiv \omega \in (\Phi_X \cup \Phi_Y)^c$:

$$\begin{aligned}
|X(\omega) - Y(\omega)| &\leq |X(\omega) - X_n(\omega)| + |X_n(\omega) - Y(\omega)| \\
&= |X(\omega) - X(\omega)| + |Y(\omega) - Y(\omega)| \qquad \text{as } n \to \infty \\
&= 0 + 0 \\
&= 0
\end{aligned}$$

Uniqueness in almost surely Convergence (2)

- We assume there are some domain $\Omega_\Phi \subseteq (\Phi_X \cup \Phi_Y)$. such that $\{\forall\, \omega \in \Omega_\Phi : X(\omega) \neq Y(\omega)\}$as $n \to \infty$:

$$\begin{aligned} P(X \neq Y) &= P(\{\omega \in \Omega_\Phi : X(\omega) \neq Y(\omega)\}) \\ &\leq P(\{\omega \in (\Phi_X \cup \Phi_Y)\}) \\ &\leq P(\{\omega \in \Phi_X\} + P\{\omega \in \Phi_Y\}) \end{aligned}$$

- Now, what is the upper-bound $P(\{\omega \in \Phi_X\}$ and $P\{\omega \in \Phi_Y\})$?

$$\begin{aligned} &\text{If } \Phi_X = \{\omega : X_n(\omega) \not\to X(\omega) \text{ as } n \to \infty\} \\ \Longrightarrow\ & \Phi_X^c = \{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} \\ \Longrightarrow\ & P(\Phi_X^c) = 1 \text{ (by defintion of almost sure convergence)} \\ \Longrightarrow\ & P(\Phi_X) = 1 - P(\Phi_X^c) = 0 \end{aligned}$$

- Likewise, $P(\Phi_Y) = 1 - P(\Phi_Y^c) = 0$
- Therefore, we proved

$$P(X \neq Y) \leq P(\{\omega \in \Phi_X\} + P\{\omega \in \Phi_Y\}) = 0 + 0 = 0$$

- We assume there are some domain $\Omega_\Phi \subseteq (\Phi_X \cup \Phi_Y)$. such that $\{\forall\ \omega \in \Omega_\Phi : X(\omega) \neq Y(\omega)\}$as $n \to \infty$:

$$
\begin{aligned}
P(X \neq Y) &= P(\{\omega \in \Omega_\Phi : X(\omega) \neq Y(\omega)\}) \\
&\leq P(\{\omega \in (\Phi_X \cup \Phi_Y)\}) \\
&\leq P(\{\omega \in \Phi_X\} + P\{\omega \in \Phi_Y\})
\end{aligned}
$$

- Now, what is the upper-bound $P(\{\omega \in \Phi_X\}$ and $P\{\omega \in \Phi_Y\})$?

$$
\begin{aligned}
&\text{If } \Phi_X = \{\omega : X_n(\omega) \not\to X(\omega) \text{ as } n \to \infty\} \\
\implies &\Phi_X^c = \{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} \\
\implies &P(\Phi_X^c) = 1 \text{ (by defintion of almost sure convergence)} \\
\implies &P(\Phi_X) = 1 - P(\Phi_X^c) = 0
\end{aligned}
$$

- Likewise, $P(\Phi_Y) = 1 - P(\Phi_Y^c) = 0$
- Therefore, we proved

$$
P(X \neq Y) \leq P(\{\omega \in \Phi_X\} + P\{\omega \in \Phi_Y\}) = 0 + 0 = 0
$$

Useful inequalities: Cauchy-Schwarz's inequality

Let X and Y be jointly distributed random variables on $\mathbb{R}$ with each having finite variance. Then:

$$(\mathbb{E}[XY])^2 \leq E[X^2]E[Y^2]$$

Woodbury matrix identity, which says:

$$\underbrace{(A+UCV)^{-1}}_{N\times N} = A^{-1} - A^{-1}U\underbrace{\left(C^{-1}+VA^{-1}U\right)^{-1}}_{M\times M}VA^{-1}$$

This is particularlly useful when we have problems such as:

$$\left(\sigma^2 I_{n\times n} + \sum_{i=1}^{M} v_i v_i^T\right)^T$$

```
N = 10;
M = 3;
sigma = 1;

U = randn(N,M); V = randn(M,N);
A = sigma * eye(N); C = randn(M,M);

answer1 = inv(A + U*C*V);
display(answer1);

answer2 = inv(A) - inv(A)*U * inv( inv(C)+ V * inv(A)*U ) * V*inv(A);
display(answer2);
```