CME308: Stochastic Methods in Engineering

Erich Trieschman

2022 Spring quarter

1 Probability

1.1 Calculus cheat sheet

Logs: $log_b(M * N) = log_b M + log_b N$ • $log_b(\frac{M}{N}) = log_b M - log_b N$ • $log_b(M^k) = klog_b M$ • $e^n e^m = e^{n+m}$
Derivatives: $(x^n)' = nx^{n-1}$ • $(e^x)' = e^x$ • $(e^{u(x)})' = u'(x)e^x$ • $(log_e(x))' = (lnx)' = \frac{1}{x}$ • $(f(g(x)))' = f'(g(x))g'(x)$
Integrals: $\int_a^b f(x)dx = \int_{g(a)}^{g(b)} f(g(u))g'(u)du$ where $g(u) = x$ • $\int_a^b u(x)v'(x)dx = u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)dx$
Infinite series and sums: $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ • $(1 + \frac{a}{n})^n \longrightarrow e^a$
$ln(1 + x) = 1 - x + \frac{x^2}{2} - \frac{x^3}{3} + \dots = \sum_{n=0}^{\infty}(-1)^n \frac{x^n}{n}$ • $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots = \sum_{n=0}^{\infty} a^x$ for $|x| < 1$

1.2 Expectation

Conditional expectation: $p_{X|Y}(x|y) = \frac{p_{x,y}(x,y)}{p_y(y)}$
Bayes theorem: $P(E_i \mid B) = \frac{P(B|E_i)P(E_i)}{\sum_{j=1}^{\infty} P(B|E_j)P(E_j)} = \frac{P(B|E_i)P(E_i)}{P(B)}$, where $E_1, E_2, \dots$ form a partition of the sample space.
Expectation: $E(X) = \sum_x xP(X = x)$, also written $E(X) = \sum_{x \in S} X(s)p(s)$, where $p(s)$ is the probability that element $s \in S$

- $E(g(X)) = \sum_i g(x_i)p_X(x_i)$ • $E(aX + b) = aE(X) + b$ • $E(X + Y) = E(X) + E(Y)$

Variance: $Var(X) = E((X - E(X)))^2) = \sigma^2$

- $Var(X) = E(X^2) - \mu^2$ • $Var(aX + b) = a^2 Var(X)$ • $Var(X + Y) = Var(X) + Var(Y)$ for X, Y independent •

Covariance: $Cov(X, Y) = E((X - E(X)(Y - E(Y)) = E(XY) - E(X)E(Y)$

- $Cov(X, X) = Var(X)$ • $Cov(aX, bY) = abCov(X, Y)$ • $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$ • $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Law of iterated expectation: $E(E(Y \mid X)) = E(Y)$
Proof: $E(Y \mid X) = \sum_y y\frac{f_{X,Y}(X,y)}{f_X(X)}$, $E(E(Y \mid X)) = \sum_x \sum_y \left(y\frac{f_{X,Y}(x,y)}{f_X(x)}\right) f_X(x) = \sum_x \sum_y yf_{X,Y}(x, y) = \sum_y yf_Y(y) = E(Y)$
Law of total probability: $P(E) = \sum_{i=-\infty}^{\infty} P(E \mid X = x)P(X)$ and $P(E) = \int_{-\infty}^{\infty} P(E \mid X = x)f(x)dx$
Variance decomposition formula: $Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X))$
Cauchy-Schwartz inequality: $E(UV)^2 \le E(U^2)E(V^2)$, with equality if $P(cU = U) = 1$ for some constant, c
Transformations of random variables: For X with density f_X and $Y = g(X)$
$F_Y(y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$ • $f_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$

1.3 Inequalities

Jensen inequality: $E(g(x)) \ge g(E(x))$ for $g(x)$ convex
Markov inequality: For $X \ge 0$, $P(X \ge t) \le \frac{E(X)}{t}$ $\forall t > 0$. **Proof:**

$$\text{Let } Y = \begin{cases} 1 & X \ge t \\ 0 & \text{otherwise} \end{cases}, \text{ Then } tY \le X \text{ since } \begin{cases} X \ge t & t * 1 \le X \\ X < t & t * 0 < X \end{cases}$$

$$tY \le X \Longrightarrow E(tY) \le E(X) \Longrightarrow tP(X \ge t) \le E(X)) \Longrightarrow P(X \ge t) \le \frac{E(X)}{t}$$

Chebyshev inequality: $P(|X - E(X)| \ge t) \le \frac{Var(X)}{t^2}$ $\forall t > 0$. **Proof:**

$$P(|X - E(X)| \ge t) = P((X - E(X))^2 \ge t^2) \le \frac{E((X - E(X))^2)}{t^2} = \frac{Var(X)}{t^2}, \text{by Markov inequality}$$

Exponential inequality: $P(X > a) \le e^{-\theta a} E(e^{\theta X})$ for all $\theta > 0$. **Proof:**

$$P(X > a) = P(\theta X > \theta a) \text{ for } \theta > 0 \Longrightarrow P(e^{\theta X} > e^{\theta a}) \le e^{-\theta a} E(e^{\theta X}), \text{ by Markov inequality}$$

(Corrolary) Upper bound on large deviations: $P(S_n < na) \le e^{-nI(x)}$. **Proof:**

$$\begin{aligned}
P(S_n > a) &\le e^{-\theta a} E(e^{\theta S_n}), \text{ by exponential inequality for } S_n = \sum_i X_i (iid) \\
&= e^{-\theta a} \prod_i E(e^{\theta X_i}) = e^{-\theta a} E(e^{\theta X_1})^n, \text{ by iid} \\
&= e^{-\theta a + n\psi(\theta)}, \text{ where } \psi(\theta) = \log E e^{\theta X_1}, \text{ the log of the MGF} \\
P(S_n > na) &= e^{-n(\theta(x)a - n\psi(\theta(x))}, \text{ minimizing RHS w.r.t } \theta \\
&= e^{-nI(x)} \text{ where } I(x) = \theta(x)a - n\psi(\theta(x))
\end{aligned}$$

1.4 Stationarity

In some cases, X_i may not be $i.i.d$, but there may still exist a statistical equilibrium:

- $\{X_1, \dots, X_n\} \stackrel{d}{=} \{X_{m+1}, \dots, X_{m+n}\}$ • $EX_1 = EX_n$ • $Cov(X_1, X_n) = Cov(X_{m+1}, X_{m+n})$, called $c(n)$ where n is lag

We can prove $c(n) \xrightarrow{p} 0$ using Chebychev and solving for $Var\overline{X}_n$ as a function of $c(n)$, leading to the bound (which goes to 0):

$$P(|\overline{X}_n - EX_1| > \epsilon) \le \frac{2}{n} \sum_i c(i)$$

1.5 Generative functions

1.5.1 Characteristic functions

- The **characteristic function** of X is $\phi_X(t) = E\exp(itX)$
- **Common characteristic functions:** $N(\mu, \sigma^2) : \exp(it\mu - \frac{1}{2}\sigma^2 t^2)$ • $Exp(\lambda) : (1 - it\lambda^{-1})^{-1}$ • $Poisson(\lambda) : \exp\lambda(e^{it} - 1)$
- **Properties:** $E[X^k] = i^{-k} E[X^k]$ • $\phi_{a_1 X_1 + \dots + a_n X_n}(t) = \phi_{X_1}(a_1 t) \dots \phi_{X_n}(A_n t)$ for X_i indep.
- The **moment-generating function** of X is $M_X(t) = E\exp(tX)$

1.6 Weak law of large numbers

For $X_1, X_2, \dots, X_n$ i.i.d. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, then for any $\epsilon > 0$

$$P(|\overline{X}_n - \mu| > \epsilon) \le \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty, \text{ by Chebyshev inequality}, \quad X_1 + \dots + X_n = S_n \approx ES_n, \text{ the "meta result"}$$

1.7 Central limit theorem

$$\sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma} \longrightarrow N(0,1) \Longleftrightarrow \sqrt{n}(\overline{X}_n - \mu) \longrightarrow N(0, \sigma^2) = \sigma N(0,1), \;\; X_i + \dots + X_n = S_n \approx N(ES_n, VarS_n), \text{ the "meta result"}$$

$$\lim_{n\to\infty} P(\frac{S_n}{\sigma\sqrt{n}} \le x) = \Phi(x), \text{ for } S_n = \sum_{i=1}^n X_i, \;\; X_1, X_2, \dots, X_n \text{i.i.d. with } E(X_i) = 0 \text{ (WLOG)}, Var(X_i) = \sigma^2$$

Proof sketch: Start with $M_{S_n}(t)$, plug in $t/(\sigma\sqrt(n))$, and use Taylor expansion to show convergence to the MGF of a normal random variable, $e^{\frac{t^2}{2}}$ **Monte Carlo:** • Sample $Y \in \mathbb{R}^d$ • Compute $X = g(Y)$ • Repeat n times • form $\overline{X}_n$ and use CLT for asymptotic behavior

1.7.1 Delta method

If g is a differentiable function at μ, $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} N(0, g'(\mu)^2\sigma^2) = g'(\mu)\sigma N(0,1)$
Proof sketch: Start with Taylor expansion $g(\bar{X}_n) \approx g(\mu) + g'(\mu)(\bar{X}_n - \mu)$ and rearrange to get $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \approx g'(\mu)\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, g'(\mu)^2\sigma(2))$. **Note:** if we find that $g'(\mu) = 0$, then repeat this process with the second derivative, $g''(\mu)$.

1.7.2 Convergence in probability

Convergence in probability: $X_n \xrightarrow{p} X$ when $P(|X_n - X| > \epsilon) \longrightarrow 0$ as $n \longrightarrow \infty$
Continuous mapping theorem: if $X_n \xrightarrow{p} X$ and g a continuous function then $g(X_n) \xrightarrow{p} g(X)$
Consistent estimator: $T_n = T_n(X_1, \ldots, X_n)$ converges in probability to $g(\theta)$, a function of the model parameter
Bounded convergence theorem: $Z_n \xrightarrow{p} Z_\infty, |Z_n| \leq c < \infty \Longrightarrow EZ_n \xrightarrow{p} EZ_\infty$
Proof starts with $|E(Z_n - Z_\infty)|$ and uses i) triangle inequality, ii) indicator functions for the case when difference is $> \epsilon, < \epsilon$
Dominated convergence theorem: $Z_n \xrightarrow{p} Z_\infty, E\beta < \infty, |Z_n(\omega)| \leq \beta(\omega) \forall \omega \Longrightarrow EZ_n \xrightarrow{p} EZ_\infty$
Fatous Lemma: for $Z_n > 0$, $E \lim_{n\to\infty} Z_n \leq \lim_{n\to\infty} EZ_n$

1.7.3 Convergence in distribution (a.k.a. weak convergence)

Convergence in distribution: $X_n \xrightarrow{d} X$ when $\lim_{n\longrightarrow\infty} F_{X_n}(x) = F_X(x)$ at all continuity points in F_X
Equalities: $X_n \xrightarrow{d} X \Longleftrightarrow Eh(Z_n) \xrightarrow{p} Eh(Z_\infty) \forall h$, bounded/continuous $\Longleftrightarrow \phi_{Z_n}(t) \xrightarrow{p} \phi_{Z_\infty}(t) \forall t$
Confidence intervals: $P(Z_{\alpha\div 2} \leq Z \leq Z_{1-\alpha\div 2}) = P(\frac{\hat{\sigma}}{\sqrt{n}} Z_{\alpha\div 2} \leq \bar{X}_n - \mu \leq \frac{\hat{\sigma}}{\sqrt{n}} Z_{1-\alpha\div 2}) = P(\mu \in \left[\bar{X}_n \pm \frac{\hat{\sigma}}{\sqrt{n}} Z_{1-\alpha\div 2}\right]) = 1 - \alpha$
Slutsky's lemma $A_n X_n + B_n \xrightarrow{d} aX + b$ if $\{X_n\}$ sequence, $X_n \xrightarrow{d} X$, $\{A_n\}$ sequence, $A_n \xrightarrow{d} A$, $\{B_n\}$ sequence, $B_n \xrightarrow{b} b$

1.7.4 Almost sure convergence

$P(\omega : \lim_{n\to\infty} X_n(\omega) \xrightarrow{p} X_\infty(\omega)) = 1$ where ω is element in set of all sequences • P($\limsup_{n\to\infty}\{|X_n - X_\infty|\} > \epsilon) = 0$

1.8 Theory of large deviations

Variance reduction: $EX = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xf(x)h(x)/h(x) = E[Zf(x)/h(x)]$ where $h(z)$ is pdf of Z. $f(x)/h(x)$ known as the likelihood ratio.
Importance sampling: Choose $h(x)$ to minimize variance. Minimal $H(dx)$ turns out to be the conditional probability of the event happening on event happening: $H^*(dx) = \mathbb{I}\{A\}(x)F(dx)/F(A)$

1.8.1 Ergotic theorem

1.8.2 Cramer-Rao Bound

1.9 Censoring data

1.10 Estimating equations

2 Statistics

2.1 Method of moments estimator

- $E(X^k) = g(\theta)$ • Calculate moment with MGF, lower moments typically lead to estimators with lower asymptotic variance
- $g^{-1}(E(X^k)) = \theta$ • Invert this expression to create an expression for the parameter(s) in terms of the moment
- $\hat{\theta} = g^{-1}(\frac{1}{n} \sum X_i^k)$ Insert the sample moment into this expression, thus obtaining estimates of the parameters in terms of data
- $\sqrt{n}(g^{-1}(\frac{1}{n} \sum X_i^k) - \theta) \xrightarrow{d} N(0, f'(E(X_i^k))^2 Var(X_i^k)^2)$ Use the delta method
- If multiple parameters characterize the distribution, use multiple moments and a system of equations

2.2 Maximum likelihood estimator

- $L(\theta) = \prod_{i=1}^n f(X_i, \theta)$ • Construct the likelihood function
- $log(L(\theta)) = l(\theta) = \sum_{i=1}^n log(f(X_i, \theta))$ • Take the log of the likelihood
- Find critical points of this function
- Find critical points of this function (e.g., $0 = \sum_{i=1}^n \frac{d}{d\theta} log(f(X_i, \hat{\theta})))$ and determine that one is a maximum

Approach to constructing MLE when indicators, $\mathbb{I}\{U\}$, are present: Logs of indicators and derivatives of indicators are very difficult to work with • Simplify likelihood function (splitting indicators when possible) • Make an argument for why the function is increasing or decreasing • Determine the value at the bounds of the function

3 Examples

3.1 Newsvendor problem