Theoretical Statistics. Lecture 2.

Peter Bartlett

1. Review: Stochastic convergence.
2. Asymptotics.
3. Concentration inequalities.

Review. Relating Convergence Properties

Theorem:

$$X_n \rightsquigarrow X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \implies Y_n \rightsquigarrow X,$$

$$X_n \rightsquigarrow X \text{ and } Y_n \rightsquigarrow c \implies (X_n, Y_n) \rightsquigarrow (X, c),$$

$$X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y \implies (X_n, Y_n) \xrightarrow{P} (X, Y).$$

Review. Relating Convergence Properties: Continuous Mapping

Suppose $f : \mathbb{R}^k \to \mathbb{R}^m$ is "almost surely continuous"
(i.e., for some S with $P(X \in S)$=1, f is continuous on S).

Theorem: [Continuous mapping]

$$X_n \rightsquigarrow X \Longrightarrow f(X_n) \rightsquigarrow f(X).$$

$$X_n \stackrel{P}{\to} X \Longrightarrow f(X_n) \stackrel{P}{\to} f(X).$$

$$X_n \stackrel{as}{\to} X \Longrightarrow f(X_n) \stackrel{as}{\to} f(X).$$

Review. Relating Convergence Properties: Slutsky's Lemma

Theorem: $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$ imply

$$X_n + Y_n \rightsquigarrow X + c,$$

$$Y_n X_n \rightsquigarrow cX,$$

$$Y_n^{-1} X_n \rightsquigarrow c^{-1} X.$$

Review. Showing Convergence in Distribution

Recall that the **characteristic function** demonstrates weak convergence:
$X_n \rightsquigarrow X \Longleftrightarrow \mathbf{E}e^{it^T X_n} \to \mathbf{E}e^{it^T X}$ for all $t \in \mathbb{R}^k$.

Theorem: [Lévy's Continuity Theorem]
If $\mathbf{E}e^{it^T X_n} \to \phi(t)$ for all t in $\mathbb{R}^k$, and $\phi : \mathbb{R}^k \to \mathbb{C}$ is continuous at 0, then $X_n \rightsquigarrow X$, where $\mathbf{E}e^{it^T X} = \phi(t)$.

Review. Uniformly tight

Definition:

X is **tight** means that for all $\epsilon > 0$ there is an M for which

$$P(\|X\| > M) < \epsilon.$$

$\{X_n\}$ is **uniformly tight** (or **bounded in probability**) means that for all $\epsilon > 0$ there is an M for which

$$\sup_n P(\|X_n\| > M) < \epsilon.$$

Review. Notation: Uniformly tight

Theorem: [Prohorov's Theorem]

1. $X_n \rightsquigarrow X$ implies $\{X_n\}$ is uniformly tight.
2. $\{X_n\}$ uniformly tight implies that for some X and some subsequence, $X_{n_j} \rightsquigarrow X$.

Review. Notation for rates: o_P, O_P

Definition:

$$X_n = o_P(1) \Longleftrightarrow X_n \xrightarrow{P} 0,$$

$$X_n = o_P(R_n) \Longleftrightarrow X_n = Y_n R_n \text{ and } Y_n = o_P(1).$$

$$X_n = O_P(1) \Longleftrightarrow X_n \text{ uniformly tight}$$

$$X_n = O_P(R_n) \Longleftrightarrow X_n = Y_n R_n \text{ and } Y_n = O_P(1).$$

Review. Relations between rates

$$o_P(1) + o_P(1) = o_P(1).$$

$$o_P(1) + O_P(1) = O_P(1).$$

$$o_P(1) O_P(1) = o_P(1).$$

$$(1 + o_P(1))^{-1} = O_P(1).$$

$$o_P(O_P(1)) = o_P(1).$$

$$X_n \xrightarrow{P} 0,\ R(h) = o(\|h\|^p) \Longrightarrow R(X_n) = o_P(\|X_n\|^p).$$

$$X_n \xrightarrow{P} 0,\ R(h) = O(\|h\|^p) \Longrightarrow R(X_n) = O_P(\|X_n\|^p).$$

Outline of the rest of today's lecture

Often we would like bounds on tail probabilities like $P(T_n \geq t)$ for some statistic T_n. We could consider asymptotic results—like the central limit theorem:

$$\lim_{n\to\infty} P(\bar{X}_n \geq \mu + \sigma\sqrt{n}t) = 1 - \Phi(t).$$

This tells us what happens asymptotically, but we usually have a fixed sample size. What can we say in that case? For example, what is

$$P(\bar{X}_n \geq \mu + \epsilon)?$$

In this lecture, we look at **deviation inequalities**, i.e., bounds on this kind of probability of deviation. We need to exploit information about the random variables.

1. Using moment bounds:
   Markov (first), Chebyshev (second)

2. Using moment generating function bounds, for sums of independent r.v.s:
   Chernoff; Hoeffding; sub-Gaussian, sub-exponential random variables; Bernstein.
3. Martingale methods:
   Hoeffding-Azuma, bounded differences.

Markov's Inequality

Theorem: For $X \geq 0$ a.s., $\mathbf{E}X < \infty$, $t > 0$:

$$P(X \geq t) \leq \frac{\mathbf{E}X}{t}.$$

Proof:

$$\begin{aligned} \mathbf{E}X &= \int X dP \\ &\geq \int_t^\infty x dP(x) \\ &\geq t \int_t^\infty dP(x) \\ &= tP(X \geq t). \end{aligned}$$

Moment Inequalities

Consider $|X - \mathbf{E}X|$ in place of X.

Theorem: For $\mathbf{E}X < \infty$, $f : [0, \infty) \to [0, \infty)$ strictly monotonic, $\mathbf{E}f(|X - \mathbf{E}X|) < \infty$, $t > 0$:

$$\begin{aligned} P(|X - \mathbf{E}X| \geq t) &= P\left(f(|X - \mathbf{E}X| \geq f(t)\right) \\ &\leq \frac{\mathbf{E}f(|X - \mathbf{E}X|)}{f(t)}. \end{aligned}$$

e.g., $f(a) = a^2$ gives **Chebyshev's inequality:**

Theorem:

$$P(|X - \mathbf{E}X| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

e.g., $f(a) = a^k$:

Theorem:

$$P(|X - \mathbf{E}X| \geq t) \leq \frac{\mathbf{E}|X - \mathbf{E}X|^k}{t^k}.$$

Chernoff bounds

Use $a \mapsto \exp(\lambda a)$ for $\lambda > 0$:

Theorem: For $\mathbf{E}X < \infty$, $\mathbf{E}\exp(\lambda(X - \mathbf{E}X)) < \infty$, $t > 0$:

$$\begin{aligned} P(X - \mathbf{E}X \geq t) &= P\left(\exp(\lambda(X - \mathbf{E}X)) \geq \exp(\lambda t)\right) \\ &\leq \frac{\mathbf{E}\exp(\lambda(X - \mathbf{E}X))}{\exp(\lambda t)}. \end{aligned}$$

$M_{X-\mu}(\lambda) = \mathbf{E}\exp(\lambda(X - \mu))$ (for $\mu = \mathbf{E}X$) is the **moment-generating function** of $X - \mu$.

Chernoff bounds

$$\begin{aligned}\log P(X-\mu \geq t) &\leq \inf_{\lambda>0}\left(-\lambda t + \log M_{X-\mu}(\lambda)\right)\\ &= -\sup_{\lambda>0}\left(\lambda t - \log M_{X-\mu}(\lambda)\right)\\ &= -\Gamma_+^*(t),\end{aligned}$$

where $\Gamma(\lambda) = \log M_{X-\mu}(\lambda)$ is the **cumulant generating function** of $X-\mu$,

$$\Gamma_+(\lambda) = \begin{cases} \log M_{X-\mu}(\lambda) & \text{if } \lambda > 0, \\ \infty & \text{otherwise,} \end{cases}$$

and Γ_+^* is the **convex conjugate** of Γ_+:

$$\Gamma_+^*(t) = \sup_{\lambda}\left(\lambda t - \Gamma_+(\lambda)\right).$$