Probability and stats mop-up unit

- Jointly distributed random variables (spring-break eve)
- Understanding covariance and correlation (last week Monday)
- Convergence of random variables (last week Wednesday)
- The weak law of large numbers (last week Friday)
- The central limit theorem (**Today**)
- Review for test (Wednesday)
- Test 2 (Friday)

Today:

- The statement of the Central Limit Theorem
- Recap of sample mean etc
- Various ways to state the CLT
- Experimental observations
- Understanding the CLT

## Theorem (The Weak Law of Large Numbers)

*If $X_0, X_1, \ldots X_{n-1}$ are a sequence of independent and identically-distributed random variables, each having mean $E[X_i] = \mu$, then*

$$\overline{X}_n \xrightarrow{\mathcal{P}} \mu$$

*that is, the sample mean converges in probability to $\mu$.*

## Theorem (The central limit theorem)

*If $X_0, X_1, \ldots X_{n-1}$ are a sequence of independent and identically-distributed random variables, each having mean $E[X_i] = \mu$ and variance $\sigma^2$, then*

$$\overline{X}_n \xrightarrow{\mathcal{D}} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{\frac{-n(x-\mu)^2}{2\sigma^2}}$$

*that is, the sample mean converges in distribution to the normal distribution with mean $\mu$ and variance $\frac{\sigma^2}{n}$.*

Let $X_0, X_1, \ldots X_n$ be random variables over the same sample space. As an ordered collection, we refer to them as a **sequence of random variables** of size $n$.

The **sample mean** of this sequence is

$$\overline{X}_n = \frac{1}{n} \sum_{i=0}^{n-1} X_i$$

If each random variable $X_i$ has the same distribution and they are independent, we refer to them as **independent and identically-distributed** (IID) random variables.

## Theorem

*If $X_0, X_1, \ldots$ be a sequence of IID random variables, $\mu = E[X_i]$, and $\sigma^2 = Var(X_i)$, then*

a. $E[\overline{X}_n] = \mu$

b. $Var(\overline{X}_n) = \frac{\sigma^2}{n}$

Note that

$$\frac{\sum_{i=0}^{n-1} X_i - n\mu}{\sigma\sqrt{n}} = \frac{n(\overline{X}_n - \mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}$$

Recall that the Gaussian distribution, with parameters $\mu$ and $\sigma^2$, is

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Moreover, the **standard Gaussian distribution** is

$$\mathcal{N}(0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

$$\begin{aligned}
\forall\ a \in \mathbb{R}, \lim_{n\to\infty} P(\overline{X}_n \leq a) &= \int_{-\infty}^{a} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{\frac{-n(x-\mu)^2}{2\sigma^2}}\, dx \\
\overline{X}_n &\approx \mathcal{N}(\mu, \tfrac{\sigma^2}{n}) \\
\overline{X}_n - \mu &\approx \mathcal{N}(0, \tfrac{\sigma^2}{n}) \\
\sqrt{n}(\overline{X}_n - \mu) &\approx \mathcal{N}(0, \sigma^2) \\
\tfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} &\approx \mathcal{N}(0, 1) \\
\forall\ a \in \mathbb{R}, \lim_{n\to\infty} P\left(\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \leq a\right) &= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx
\end{aligned}$$

Let $X_0, X_1, \ldots$ be a sequence of random variables and let $X$ be another random variable. Let $F_n$ be the cumulative distribution function of $X_n$ and $F$ be the cumulative distribution function of $X$.

$X_n$ **converges** to $X$ **in probability**, written as $X_n \xrightarrow{\mathcal{P}} X$, if for any $\epsilon \in \mathbb{R}^+$,

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0$$

$X_n$ **converges** to $X$ **in distribution**, written as $X_n \xrightarrow{\mathcal{D}} X$, if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

$X_n$ **converges** to $X$ **in quadratic mean**, written as $X_n \xrightarrow{\mathcal{QM}} X$, if

$$\lim_{n \to \infty} E[(X_n - X)^2] = 0$$

$X_n$ **converges almost surely** to $X$, written as $X_n \xrightarrow{\mathcal{AS}} X$, if

$$P(\lim_{n \to \infty} X_n = X) = 1$$

**For next time:**

*Look over review sheet*