Stochastic Calculus, Courant Institute, Fall 2012

http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2012/index.html

Always check the class message board on the blackboard site from home.nyu.edu before doing any work on the assignment.

Assignment 1, due September 17

Corrections: (Sept. 11: $\text{var}(X^4)$ fixed to be $\text{var}(X^2)$ in (1b), $Y = X\cdots$ corrected to $Y = X\cdots$ in (2(d)i), equation (4) fixed. Sept. 12: part (3b) has the inequality $|Z| \leq \cdots$ fixed to put $\sqrt{n-1}$ in the denominator.)

Many of the exercises this week are very standard. You will easily find the solutions in textbooks. I ask you not to do that, at least not before trying them on your own. You will understand the material much better if you work it out yourself.

1. (*An aspect of the central limit theorem*) Suppose Y_i are i.i.d. random variables with $E[Y_i] = 0$, $E\left[Y_i^2\right] = \sigma^2$, and $E\left[Y_i^4\right] = \mu_4 < \infty$. Suppose $X_n = \frac{1}{\sqrt{n}}\left(Y_1 + \cdots + Y_n\right)$. The CLT says that as $n \to \infty$, the distribution of X_n converges to $\mathcal{N}(0, \sigma^2)$. This means that if $V(x)$ is a reasonable function of x (what is "reasonable" depends on the application), then $E[V(X_n)]$ depends mostly on σ^2 and very little on other details of the distribution of Y_i. This exercise checks that for $V(x) = x^4$.

   (a) Show that if $X \sim \mathcal{N}(0, \sigma^2)$, then
   $$E\big[X^4\big] = 3\sigma^4 \; . \qquad (1)$$
   Hint: Write the integral formula for the expectation, use $xe^{-x^2/2\sigma^2} = **\,\partial_x e^{-x^2/2\sigma^2}$, and integrate by parts.

   (b) Conclude that the variance of X^2 is $\text{var}\big(X^2\big) = 2\sigma^4$. We will use this formula many times this semester.

   (c) Write the formula for $E\big[X_n^4\big]$. *Hint:* Use
   $$X_n^4 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n Y_iY_jY_kY_l \; ,$$
   Take the expectation and figure out which of the terms are different from zero. Among those are terms like $E\big[Y_i^2Y_j^2\big] = \sigma^4$ if $i \neq j$, and terms like $E\big[Y_i^4\big] = \mu_4$. You need to figure out how many terms of each kind there are.

   (d) From your formula for part 1c, show that $E\big[X_n^4\big] \to 3\sigma^4$ as $n \to \infty$. This does not have to be a formal mathematical proof, just use the fact that $\frac{1}{n} \to 0$ as $n \to \infty$. The CLT says that the influence of μ_4 should disappear in the limit $n \to \infty$.

(e) (*not an action item*) The non-zero terms in the sum representing $E\left[X_n^4\right]$ are $\frac{\sigma^4}{n^2}$ or $\frac{\mu_4}{n^2}$. They have the same order of magnitude in a mathematical sense because they share the same power of n. You cannot tell how important terms of a certain kind are by looking at just one term. You have to know more about an expansion to know which terms add up to something significant and which do not.

2. (*Student $t-$distribution, part 1*) The Student $t-$distribution is important in statistics because it is related to the accuracy of estimates of the mean of a Gaussian random variable. It is becoming widely used also because it is a simple probability distribution with an explicit PDF that has power-law fat tails depending on a parameter. Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ is a one dimensional Gaussian. Suppose X_i are independent samples of X. An estimator of the distribution mean is the sample mean:

$$\widehat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \ . \tag{2}$$

An estimator of the variance is

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 \ . \tag{3}$$

Both $\widehat{\mu}$ and $\widehat{\sigma^2}$ are random variables. This exercise studies their distribution.

(a) Show that if X_i is replaced by $X_i - \mu$, then the distribution of $\widehat{\sigma^2}$ is unchanged and the distribution of $\widehat{\mu}$ is only shifted by μ. This is mainly theoretical because you are unlikely to know μ in practice. The distribution of $X_i - \mu$ is $\mathcal{N}(0, \sigma^2)$. From now on, assume that $\mu = 0$.

(b) If X_i is replaced by $\frac{1}{\sigma} X_i$, how do the random variables $\widehat{\mu}$ and $\widehat{\sigma^2}$ change? From now on, assume $\sigma = 1$, so $X_i \sim \mathcal{N}(0,1)$.

(c) Let $v_1 \in \mathbb{R}^n$ be the column vector $v_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)^t$. Find $\|v_1\|_{l^2}$. Show that there are vectors v_2, $\ldots$, v_n so that the vectors v_1, v_2, $\ldots$, v_n are an ortho-normal basis of $\mathbb{R}^n$. You can do the second part abstractly or by quoting a theorem (if you state it completely).

(d) Let $v \in \mathbb{R}^n$ be any unit vector ($\|v\| = 1$, $\|v\| = \|v\|_{l^2}$ everywhere in this exercise). Let $\mathcal{S} \subset \mathbb{R}^n$ be the plane of vectors perpendicular to v. That is, $x \in \mathcal{S}$ if and only if $x^t v = 0$. The *orthogonal projection* of x onto $\mathcal{S}$ is the $y \in \mathcal{S}$ that minimizes $\|x - y\|$. Show the basic facts about projections, not necessarily in this order:

i. $y = x - (x^t v)\, v$ $\qquad$ ($(x^t v)$ is the v component of x.)

ii. y is perpendicular to $x - y$ $\qquad$ (the geometry of $l^1 2$ projection)

iii. $\|x\|^2 = \|y\|^2 + (x^t v)^2$ $\qquad$ (the Pythagorean theorem)

(e) From now on, $\mathcal{S}$ is the plane perpendicular to v_1 of part (2c). Show that $\overline{X} = \frac{1}{\sqrt{n}} X^t v_1$.

(f) Let $Y_i = X^t v_i$ for $i = 2, \ldots, n$. Show that the Y_i are independent of each other, independent of $\overline{X}$ and have $Y_i \sim \mathcal{N}(0,1)$.

(g) Let Y be the orthogonal projection of X onto $\mathcal{S}$. Show that

$$\|Y\|^2 = \sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} \left(X_i - \left[X^t v_1\right] v_{1,i}\right)^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

(h) Show that the random variables $\widehat{\sigma}^2$ and $\widehat{\mu}$ are independent.

(i) The *chi square* distribution with k *degrees of freedom* is defined as the distribution of $Q = Z_1^2 + \cdots + Z_k^2$, where $Z_i \sim \mathcal{N}(0,1)$ are independent. This is written $Q \sim \chi_k^2$. Show that $\widehat{\sigma^2} \sim \frac{1}{n-1}\chi_{n-1}^2$ and that $E\left[\widehat{\sigma^2}\right] = 1$.

(j) Return now to $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \neq 0$ and $\sigma^2 \neq 1$. Then $\overline{X}$ is normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. Measuring $\widehat{\mu} - \mu$ in units of its estimated standard deviation gives rise to the ratio

$$t = \frac{\sqrt{n}\,(\widehat{\mu} - \mu)}{\sqrt{\widehat{\sigma^2}}} \; . \tag{4}$$

Show that t has the same distribution as

$$t = \frac{\sqrt{n-1} Z}{\sqrt{\chi_{n-1}^2}} \; , \tag{5}$$

where Z is a one dimensional standard normal and χ_{n-1}^2 is an independent chi-square random variable with $n-1$ degrees of freedom. Conclude that the t in (4) is independent of the parameters μ and σ. This independence is the basis of the *Student* $t-$*test* in statistics.

3. Part (2) is all you need to know about the $t-$statistic to do statistics. But the formula for the density of the t random variable is useful in many other modeling problems. In this problem, C represents a normalization constant that may be different in different places.

(a) Let $Q \sim \chi_{n-1}^2$. Let $F(Q)$ be the distribution function $F(q) = \Pr(Q \leq q)$. Let $f(q) = F'(q)$ be the probability density. Show that

$$F(q) = C \int_{\|x\|^2 \leq q} e^{-\|x\|^2/2}\, dx = C \int_{r=0}^{\sqrt{q}} r^{n-1} e^{-r^2/2}\, dr \; ,$$

and that

$$f(q) = C q^{(n-2)/2} e^{-q/2} \; . \tag{6}$$

It is possible to evaluate the constant explicitly: $C = \Gamma(n/2)2^n$, but we will not need this formula here and I am not asking you to derive it. In fact, it's essentially the definition of Γ. Since the full formula for $f(q)$ involves a Gamma function, the distribution is sometimes called a Gamma distribution.

(b) We want the density of t defined by (5). This is clearly symmetric: $\Pr(t = -x) = \Pr(t = x)$. Let $f(x)$ be that density, defined by $f(x)dx = \Pr(x \leq t \leq x + dx)$. Since $f(-x) = f(x)$, we can find f from a modified distribution function $F(x) = \Pr(|t| \leq x)$, which has $f(x) = 2F'(x)$. Show that $|t| \leq x$ is equivalent to $|Z| \leq x\sqrt{Q}/\sqrt{n-1}$. Write the z integral that represents that probability for fixed Q, then write the double integral over z and q that represents that probability. The variable x appears only in the limit of integration of the inner (dz) integral.

(c) Differentiate this expression with respect to x to get a one dimensional integral expression of the form

$$f(x) = C \int_{q=0}^{\infty} q^p e^{-a(x)q}\, dq \ ,$$

with $a(x)$ and p explicitly given as simple functions of n and x.

(d) Use the change of integration variable $a(x)q = r$ to get an explicit formula

$$f(x) = \frac{C}{a(x)^{p+1}} \ .$$

If you did all this correctly, the answer should be

$$f(x) = \frac{C}{\left(1 + \frac{1}{n-1}x^2\right)^{n/2}} \ .$$

This is the *Student $t-$density* with $n-1$ degrees of freedom. More generally, the $t-$density with *parameters* μ and σ and n degrees of freedom is

$$f(x;\mu,\sigma,n) = \frac{C}{\left(1 + \dfrac{(x-\mu)^2}{n\sigma^2}\right)^{n/2+1}} \ . \qquad (7)$$

There is an explicit formula for C, but it is bigger than the rest of the expression and rarely is needed. An important feature of this formula is that n does not have to be an integer. Of course, it was an integer in (5), but even that is unnecessary if we use (6) to define the chi-square distribution for non-integer n.

(e) Clearly $\mu = E[X]$ for the density (7). Show that $\mathrm{var}(X) < \infty$ for $n > 1$ but the formula $\sigma^2 = \mathrm{var}(X)$ is not true. *Hint #1*: It suffices to show $E\big[X^2\big] \neq 1$ when $\mu = 0$ and $\sigma = 1$, and for some value of

n. *Hint #2*: This may be very time consuming. Please do not do it unless you have lots of free time and everything else is finished. The fact is more important than the proof.

(f) Show that the probability density (7) has a *power law tail*, which means that $f(x) \approx Cx^{-p}$ as $x \to \pm\infty$.

4. Suppose that $Z_n \sim \mathcal{N}(0,1)$ and

$$X_{n+1} = \frac{1}{2}X_n - \frac{5}{16}X_{n-1} + Z_n \; .$$

Show that this Gaussian process is stable. Find the limiting joint distribution of X_n and X_{n-1} in the limit $n \to \infty$.