Stochastic Calculus, Courant Institute, Fall 2019

http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2019/StochCalc.html

Always check the `classes` message board before doing any work on the assignment.

Assignment 1, due September 16

Corrections: 9/11: standard deviation formula from question 2 corrected to $\frac{1}{\sqrt{m\Delta x}}\sqrt{\widehat{p}_k}$

1. (*About histograms*) Suppose $p(x)$ is the PDF for a one-component random variable X. Assume that $X_j \sim p$ is a collection of i.i.d. samples of p. Suppose there is a *bin size*, Δx, and bin k is $B_k = (a+k\Delta x, a+(k+1)\Delta x)$. The *bin centers* are $x_k = a + (k+\frac{1}{2})\Delta x$. The *bin counts* with m samples are the random numbers $N_k = \#\{X_j \in B_k \mid j = 1, \ldots, m\}$. Define the *scaled* bin counts by

$$\widehat{p}_k = \frac{1}{\Delta x\, m} N_k \ .$$

A *histogram* is a plot of N_k or $\widehat{p}_k$ as a function of k or x_k. Show that (if $p'(x)$ and $p''(x)$ are continuous functions of x)

$$\mathrm{E}[\,\widehat{p}_k] = p(x_k) + O(\Delta x^2) \ .$$

[This exercise is partly for the result: the scaled expected bin count estimates the probability density. It's also for the "analytical technique", the "big Oh" notation and how it's justified. If $Q(s)$ is defined for $s > 0$ and there is a fixed C so that $|Q| < Cs$, then we say Q is "of the order of s", and we write $Q = O(s)$. Here, Q is $\mathrm{E}[\cdot] = p(x_k)$ and s is Δx^2. A "Taylor approximation with remainder" theorem from Calculus I (if they taught it like this) is $f(y) = f(x) + (y-x)f'(x) + \frac{1}{2}(y-x)^2 f''(\xi)$, where ξ is some number between y and x. If g is continuous, then (another theorem) there is a D with $|g(\xi)| \leq D$ for all ξ in any interval. Apply this with $g = f''$. The C is related to D and is found by integrating the Taylor inequality over B_k with respect to y.]

2. (*histogram error bar*). Suppose you estimate Q as the average of m i.i.d. samples:

$$Q \approx \widehat{Q} = \frac{1}{m}\sum_{j=1}^{m} U_j \ .$$

Assuming $Q = \mathrm{E}[U_j]$, the *statistical error* is roughly the size of the standard devation

$$\text{std. dev.}(\widehat{Q}) = \sqrt{\mathrm{var}(Q)} = \frac{1}{\sqrt{m}}\sqrt{\sigma_U^2} \ .$$

For the histogram, let $U_j = 1$ if $X_j \in B_k$ and $U_j = 0$ otherwise. This is a *Bernoulli* random variable. Explain why $\sigma_U^2 \approx p(x_k)\Delta x$ is an accurate approximation for small Δx. Use this to explain why

$$\text{std. dev.}(\widehat{p}_k) \approx \frac{1}{\sqrt{m\Delta x}}\sqrt{\widehat{p}_k}$$

is an accurate approximation if Δx is small and m is large enough. If you estimate $p(x_k)$ using $\widehat{p}_k$ (this is the histogram estimate of the probability density), the same data gives an *error bar*, which is an estimate of the size of the error in the statistical estimate.

3. *Concentration theorems* say that some random variable that depends on n is *concentrated* when n is large. A random variable S is concentrated if it is unlikely to be far from its mean. The random variables S_n are concentrated if the variation of S_n goes to zero as n goes to infinity in some sense. This is not a precise definition because there are different concentration theorems.

   Let X_k be a family of independent Gaussian random variables with mean zero and variance 1. This exercise is about the magnitude of the vector $X = (X_1, \ldots, X_n)$, which is the random variable

$$R = |X| = \left(\sum_{k=1}^{n} X_k^2\right)^{\frac{1}{2}} .$$

   Since R^2 is the sum of independent random variables with mean $\mathrm{E}\big[\,X_k^2\big] = 1$, we have $\mathrm{E}\big[\,R^2\big] = n$. Therefore, if $R = \sqrt{R^2}$ is concentrated, it should be concentrated near $\sqrt{n}$.

   Let $p_n(r)$ be the probability density of R. Show that

$$p_n(r) = \frac{1}{Z} r^{n-1} e^{-\frac{r^2}{2}} ,$$

   where

$$Z = \int_0^\infty r^{n-1} e^{-\frac{r^2}{2}}\, dr .$$

   It is common in applied probability that you know the functional form of a probability density (e.g., $r^{n-1}e^{-\frac{r^2}{2}}$) but not the *normalization constant*, Z. *Hint,* $p_n(r)dr$ is the probability that $r \leq |X| \leq r + dr$. This is equal to $e^{-\frac{r^2}{2}} A_n(r)dr$, where $A_n(r)$ is the "area" of the sphere in n dimensions given by $|x| = r$. Since $A_n(r)$ represents an $n-1$ dimensional "area", it "scales like" r^{n-1}, which a way of saying that $A_n(r) = C_n r^{n-1}$. You can find a formula for C_n in books (where it's probably called ω_{n-1}), but that formula may not be so useful. If you're not convinced by the area argument, let $V_n(r)$ be the n dimensional volume of the ball $|x| \leq r$. "Clearly" $V_n(r + dr) - V_n(r) = A_n(r)dr$, which is the same as saying

$A_n(r) = V_n'(r)$. The volume scales as r^n, i.e., $V_n(r) = D_n r^n$ (D_n being the volume $V_n(1)$, for example). The point of this task is that the power r^{n-1} can be found by "easy" scaling arguments while the constant Z is harder. If you can't evaluate Z explicitly, it's easy to compute numerically.

4. If $f(t)$ is a differentiable function of t, then the *total variation* between 0 and T is

$$\mathrm{TV}(f) = \int_0^T |f'(t)| \; dt \; .$$

This measures how much f "moves". Let $\Delta t > 0$ be a small "time step" parameter and define discrete times $t_k = k\Delta t$. An approximate total variation is

$$\mathrm{TV}(f, \Delta t) = \sum_{t_k < T} |f(t_{k+1}) - f(t_k)| \; .$$

(a) Show that if f "nice" (has continuous first and second derivatives, say), then $\mathrm{TV}(f, \Delta t) \to \mathrm{TV}(f)$ as $\Delta t \to 0$.

(b) Let X_t be a standard Brownian motion. Show that

$$\mathrm{E}[TV(X, \Delta t)] \approx \frac{C_T}{\sqrt{\Delta t}} \to \infty \; ,$$

as $\Delta t \to 0$. Evaluate C_T. *Hint,* use the independent increments property and the fact that $X_{t_{k+1}} - X_{t_k}$ is Gaussian with mean zero and known variance.

Part b suggests that the total variation of Brownian motion is infinite. We will see that this is true.

Computing assignment

Task 1. Download the Python code `IntegrationDemo.py` and run it. You should get $Z \approx 2$. Check that this is the right answer. Check that the code also gives the right answer for $n = 4$.

Task 2. Download the Python code `HistogramDemo.py` and the picture `HistogramDemoCheck.pdf`. Run the code and see that the picture it makes is the same as the picture you downloaded. Note that the error bar is estimated as in exercise 2 above. Show that if Δx is too small for a given m then the density estimate is poor. Choose Δx and m so that the error is not visible in the plot. First you need to take Δx so small that the integration error from exercise 1 is not visible. Then you need to take m so large that the error bar cannot be seen either. Note that the code uses `n` for the variable we call m.

Task 3. Modify `HistogramDemo.py` so that the random variable is called R (instead of X) and the number of samples is called m. Add a parameter n to the code and write some code that generates a sample R by generating n independent standard normals and computes R as above. Plot a histogram and the density calculated from exercise 3 with Z calculated from `IntegrationDemo.py`. If you do everything right, the histogram will fit the theoretical density. Modify the plot so that n appears in the plot title.