Homework #3

36-754, Spring 2007

Due 14 May 2007

1 Convergence of the empirical CDF, uniform samples

In this problem and the next, X_i are IID samples on the real line, with cumulative distribution function F. The empirical distribution function $\hat{F}_n$ is

$$\hat{F}_n(t) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,t])}(X_i)$$

and the empirical process is

$$U_n(t) = \sqrt{n}[\hat{F}_n(t) - F(t)]$$

where t runs over the whole range of X. As n goes to infinity, U_n converges in distribution to a Gaussian process which depends only on F, and not on n. Consequently, $\hat{F}_n$ is in some sense converging on F at rate $n^{-1/2}$. This "empirical central limit theorem" is made precise in this problem and the next.

For the rest of this problem, we consider the special case where the X_i are uniformly distributed on the unit interval, so that $F(t) = t$, $0 \leq t \leq 1$.

Here the limiting form of U_n is what's generally called the *Brownian bridge*, a real-valued process on $[0, 1]$ defined by

$$B(t) = W(t) - tW(1)$$

where W is a standard Wiener process.

1. Show that $\mathbf{E}[B(t)] = 0$ for all t, and that $\text{cov}(B(s), B(t)) = s(1-t)$ when $0 \leq s \leq t \leq 1$.
2. Show that the Brownian bridge is a Gaussian process.
3. Show that, for all n and all $0 \leq s \leq t \leq 1$, $\mathbf{E}[U_n(t)] = 0$, $\text{cov}(U_n(s), U_n(t)) = s(1-t)$.
4. Show that $U_n \stackrel{fd}{\rightarrow} B$. One way is to use the multivariate central limit theorem.

5. Show that $U_n \xrightarrow{d} B$ (in the Skorokohod sense). One way is to use Proposition 201 from §15.1. *Hint*: think about Chebyshev's inequality, as in §16.2.2. This is the "empirical central limit theorem" for the uniform distribution.

2 Convergence of the empirical CDF, continued

Now suppose that X_i have a non-uniform distribution, with cumulative distribution function F. Define the *quantile function*

$$q_F(p) \equiv \sup x : \; F(x) < p$$

1. Let Z_i be IID uniform. Show that $q(Z_i) \stackrel{d}{=} X_i$.

2. Define a map $H : \mathbf{D}([0,1]) \mapsto \mathbf{D}(\mathbb{R}^+)$ by $Hw(x) = w(F(x))$. Show that

$$\sup_x |Hw(x) - Hv(x)| \leq \sup_x |w(x) - v(x)|$$

   so that H is uniformly continuous.

3. Let $B_F = HB$, where B is the Brownian bridge. Show that $\mathbf{E}[B_F(t)] = 0$ for all t, that $\operatorname{cov}(B_F(s), B_F(t)) = F(s)(1 - F(t))$, and that B_F is a Gaussian process.

4. Show that $U_n \xrightarrow{d} B_F$. One approach is to use the previous problem and the continuous mapping theorem.

5. Using the result of exercise 16.2 in Kallenberg, show that, for each $\epsilon > 0$,

$$\mathbb{P}\left(\sup_t |\hat{F}_n(t) - F(t)| > \epsilon\right) \to 0$$

   as $n \to \infty$.

6. The result of the previous part is similar to, but not the same as, the Glivenko-Cantelli Theorem, which asserts that

$$\sup_t |\hat{F}_n(t) - F(t)| \to 0$$

   almost surely, as $n \to \infty$ (see e.g. Proposition 4.24 on p. 75 of Kallenberg). However, it does much of the same work as the Glivenko-Cantelli Theorem, by establishing that the empirical distribution function is "probably approximately correct". Explain exactly how the Glivenko-Cantelli Theorem is stronger than the convergence-in-probability result you have just proved; that is, what behavior for $\hat{F}_n$ is possible according to your result but forbidden by Glivenko-Cantelli?

7. *Extra credit*: do exercise 16.2 in Kallenberg.

3 Simulation and p-Values

1. The functional central limit theorem tells us that rescaled random walks Y_n converge in distribution on the Wiener process W. Show that the transformed rescaled walks, $B_n(t) = Y_n(t) - tY_n(1)$, converge in distribution on the Brownian bridge $B(t) = W(t) - tW(1)$.

2. Write a program to simulate the Brownian bridge. (One approach is to use the previous part, but if you can think of another way to do this, feel free to use it, with an argument for why it works.)

3. Use your simulator to get an approximate distribution for $B^* \equiv \sup_{t\in[0,1]} |B(t)|$. Can you identify the form of this distribution?

4. Write a program to simulate taking n samples from a uniform distribution on the unit interval, and to calculate $D_n = \sup_{t\in[0,1]} \sqrt{n}|\hat{F}_n(t) - t|$.

5. Using the second simulator, approximate the distribution of D_n for $n = 10, 100, 1000$ and $10,000$. Compare these distributions to the one you got for B^*. Are they converging? Should they be?

6. *Extra credit*: There is an analytical expression — a convergent series with infinitely many terms — for the cumulative distribution function of B^*; find it. (*Hints*: Express the event $B^* \leq a$ in terms of the underlying Wiener process, and use the principle of inclusion-exclusion. This one is hard!)

4 "The square of dW is dt"

Let W be a standard Wiener process, and t an arbitrary positive real number. For each n, let $t_i = it2^{-n}$.

1. Show that $\sum_i (\Delta W(t_i))^2$ converges on t (in L_2) as n grows. *Hint:* Show that the terms in the sum are IID, and that their variance shrinks sufficiently fast as n grows. (You will need the fourth moment of a Gaussian distribution.)

2. If $X(t)$ is measurable, non-anticipating and square-integerable, show that

$$\lim_n \sum_{i=0}^{2^n-1} X(t_i)(\Delta W(t_i))^2 = \int_0^t X(s)ds \quad (1)$$

in L_2.

5 Diffusion approximation to a branching process

In a *branching process*, we model the growth or decay of a population of identical objects, conventionally called "particles". (These models can be applied in genetics, ecology, chemistry, astrobiology and, most spectacularly, nuclear engineering.) Particles persist for a random length of time, and at each time step during which a particle is around, it produces a random, possibly zero, number of further particles ("offspring"), independently of what other particles are doing and of what it did at other times. That is, if we write X_n for the total number of particles at the n^{th} time-step, then

$$X_{n+1} = \sum_{i=1}^{X_n} Y_{in}$$

where the $Y_{in} \geq 0$ are all independent and identically distributed random integers. (If $Y_{in} = 0$, then particle i died at time-step n without offspring. If $Y_{in} = 1$, then either i lived without offspring, or it died after producing one child — since all particles are identical it doesn't matter which.) Write $\mathbf{E}[Y_{in}] = \mu$, $\mathbf{Var}[Y_{in}] = \sigma^2$, both finite.

Throughout, assume the branching process always begins with a single particle, $X_0 = 1$.

1. Show that X is a Markov chain, regardless of the distribution of the Y_{in}.
2. Show that
$$\mathbf{E}[X_{n+1} - X_n | X_n = x] = (\mu - 1)x$$
and that
$$\mathbf{Var}[X_{n+1} - X_n | X_n = x] = \sigma^2 x$$
3. Find the time-evolution operator K, i.e.,
$$\mathbf{E}[f(X_{n+1}) | X_n = x] = Kf(x)$$
For what functions f is this well-defined?
4. Consider a sequence of branching processes $X^{(k)}$ where $\mu_k = 1 + \mu/k$ and $\sigma_k^2 = \sigma^2/k$. Define a continuous-time process $Z^{(k)}(t)$ by
$$Z^{(k)}(t) = \frac{X^{(k)}_{\lfloor kt \rfloor}}{k}$$
Show that, as $k \to \infty$, $Z^{(k)}(t)$ converges in distribution on a Gaussian random variable, and find the mean and variance in terms of μ and σ^2.
5. Show that $Z^{(k)} \xrightarrow{d} Z$, a Feller process whose generator G is
$$c_1 x \frac{\partial}{\partial x} + c_2 x \frac{\partial^2}{\partial x^2}$$
and find c_1 and c_2 in terms of μ and σ^2.

6 The White Noise Model and Non-Parametric Regression

Let f and σ be two continuous, measurable real functions on the unit interval. When the independent variable is $x \in [0,1]$, we observe $Y = f(x) + \epsilon$, where ϵ are IID noise terms with mean 0 and variance $\sigma^2(x)$. (That is, the noise is heteroskedastic.) This problem relates estimating the regression function f to estimating the coefficients of a small-noise SDE. For each positive integer n, $x_{ni} = i/n$, $0 \leq i < n$. Then $Y_{ni} = f(x_{ni}) + \epsilon_{ni}$, $\epsilon_{ni} \sim \mathcal{N}(0, \sigma^2(x_{ni}))$. Define the continuous-parameter process

$$R_n(t) \equiv \frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor} Y_{ni}$$

1. Show that, for each fixed t,

$$R_n(t) \xrightarrow{d} \int_0^t f(s)ds + \frac{1}{\sqrt{n}}\int_0^t \sigma(s)dW$$

2. Define Z_n as the solution to the SDE $dZ_n = f(t)dt + n^{-1/2}\sigma(t)dW$, with initial condition $Z_n(0) = 0$. Show that each Z_n is a Gaussian diffusion.

3. The previous two parts suggest that R_n and Z_n should be converging. One form of this convergence would be

$$\mathbb{P}\left(\sup_{t\in[0,1]} |R_n(t) - Z_n(t)| > \delta\right) \to 0$$

as $n \to \infty$. Prove this, or find a counter-example.

4. Let ϕ_i, $i = 1, 2, \ldots$ be an orthonormal basis for $L_2([0,1])$, i.e.,

$$\int \phi_i(x)\phi_j(x)dx = \delta_{ij}$$

For each ϕ_i, define

$$\theta_i = \int f(x)\phi_i(x)dx$$

Show that

$$\theta_i = \int \phi_i(t)dZ_n(t) + n^{-1/2}\eta_i$$

where η_i is a mean-zero Gaussian random variable. Your proof should also include an expression for the variance of η_i, and show that it does not change with n. Are the η_i independent? (*Hint:* Use Itô's isometry.)

This correspondence between non-parametric regression and white-noise problems can be extended to include the case where the design points x_{ni} are not evenly spaced, and in fact to where they are randomly positioned.