Solutions to Problem Set 2

Edited by Chris H. Rycroft*

October 5, 2006

1 Asymptotics of Percentile Order Statistics

Let X_i ($i = 1, \ldots, N$) be independently identically distributed (IID) continuous random variables with CDF, $P(x) = \mathbb{P}(X_i \leq x)$, and PDF, $p(x) = \frac{dP}{dx}$. Let $Y_N^{(\alpha)}$ be the 100αth percentile, which is uniquely defined by ordering the outcomes, $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(N)}$ (using the standard notation of *order statistics*) and setting $Y_N^{(\alpha)} = X_{([\alpha N])}$ (where $[x]$ is the nearest integer to x).

1.1 Expression for the PDF of $Y_N^{(\alpha)}$

To derive the PDF $f_N(y)$ of $Y_N^{(\alpha)}$, we calculate the probability that the random variable $Y_N^{(\alpha)}$ lies in an interval $(y, y + dy)$, which equals $f_N(y)dy$ in the limit $dy \to 0$. Thinking of the event $\left\{y < Y_N^{(\alpha)} = X_{([\alpha N])} < y + dy\right\}$ in terms of the set of ordered outcomes

$$\underbrace{X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{([\alpha N]-1)}}_{([\alpha N]-1)} \leq X_{([\alpha N])} \leq \underbrace{X_{([\alpha N]+1)} \leq \ldots \leq X_{(N)}}_{(N-[\alpha N])}$$

we see that the event we are interested in can be broken down into three steps:

1. choose $X_{([\alpha N])}$ between y and $y + dy$;
2. choose $([\alpha N] - 1)$ of the X_i's less than y;
3. choose $(N - [\alpha N])$ of the X_i's greater than $y + dy$.

Therefore, we find that

$$\mathbb{P}\left(y < Y_N^{(\alpha)} < y + dy\right) = N\binom{N-1}{[\alpha N]-1} P(y)^{[\alpha N]-1} \left(1 - P(y)\right)^{N-[\alpha N]} p(y)dy,$$

where the factor $N\binom{N-1}{[\alpha N]-1}$ counts the number of different ways that we could have selected the X_i. The PDF $f_N(y)$ is simply this probability divided by dy:

$$f_N(y) = N\binom{N-1}{[\alpha N]-1} P(y)^{[\alpha N]-1} \left(1 - P(y)\right)^{N-[\alpha N]} p(y). \tag{1}$$

*Based on solutions for problems 2 and 4 by Chris H. Rycroft (2005), for problem 1 by Kevin Chu (2003), and for problem 3 by Michael Slutsky (2003).

Note that unlike the random variable $Z_N = \frac{1}{N}\sum_{n=0}^{N} X_n$, $Y_N^{(\alpha)}$ does not necessarily tend towards the central regions of the parent PDF $p(x)$ as $N \to \infty$. For example, for $\alpha = 1$ and N large, $Y_N^{(\alpha)}$ is most likely to be found in the upper tail of $p(x)$ because the probability that all of the X_i's are less than some y is given by $P(y)^N$ which is very small if y is not in the upper tail.

Intuitively, the reason that $Y_N^{(\alpha)}$ does not tend towards the central region, is that as $N \to \infty$, the histogram of the X_i's should tend towards the parent PDF $p(x)$. As a result, we would expect that $Y_N^{(\alpha)}$ to tend towards the value x_α which satisfies $P(x_\alpha) = \alpha$, which is only the "central region" for $\alpha \approx P(x_{\text{mean}})$.

1.2 Asymptotic Approximation for the PDF of $Y_N^{(\alpha)}$

Notice that $f_N(y)$ bears some resemblance to the probability that an asymmetric Bernoulli random walk takes $[\alpha N] - 1$ steps to the right out of a total of $N-1$. This suggests that we should be able use the results from the asymptotic of a symmetric Bernoulli random walk to find a globally-valid asymptotic approximation for $f_N(y)$ as $N \to \infty$.

While it is certainly possible to carry out the asymptotic analysis directly on the asymmetric Bernoulli random walk, a lot of work can be saved by first considering the discrete walk. Let p be the probability for the walker to take a step to the right and $q = 1-p$ be the probability for the walker to take a step to the left. Then, the exact probability that the walker takes m right steps out of a total n steps is

$$P_n(2m-n) = \binom{n}{m} p^m q^{n-m}$$

where $P_n(i)$ is the probability that the walker is at position i after n steps. This expression can be conveniently rewritten in terms of the probability for a symmetric Bernoulli walk:

$$P_n(2m-n) = 2^n p^m q^{n-m} \left[\binom{n}{m} \left(\frac{1}{2}\right)^n \right].$$

Combining this result and the globally-valid asymptotic approximation for the symmetric Bernoulli walk, we see that

$$\begin{aligned} P_n(2m-n) &= 2^n p^m q^{n-m} P_{\text{sym. Bernoulli}}(2m-n) \\ &\sim 2^n p^m q^{n-m} \left[\frac{2}{\sqrt{2\pi n(1-\xi^2)}} e^{-n\psi(k_s)} \right] \end{aligned}$$

where

$$\begin{aligned} \psi(k_s) &= \frac{\xi}{2}\log\left(\frac{1+\xi}{1-\xi}\right) + \frac{1}{2}\log\left(1-\xi^2\right) \\ \xi &= \frac{2m-n}{n} = \frac{2m}{n} - 1. \end{aligned}$$

Using this approximation in the expression for $f_N(y)$, we find that

$$f_N(y) \sim Np(y) 2^N P(y)^{[\alpha N]-1} \left(1 - P(y)\right)^{N-[\alpha N]} \left[\frac{1}{\sqrt{2\pi(N-1)(1-\xi^2)}} e^{-(N-1)\psi(k_s)} \right]$$

with

$$\xi = \frac{2([\alpha N]-1)}{N-1} - 1.$$

Taking the continuum limit, $N \to \infty$, we replace $[\alpha N]$ by αN. In this limit, $\xi \to 2\alpha - 1$, $1-\xi^2 \to 4\alpha(1-\alpha)$ and

$$\begin{aligned}(N-1)\psi(k_s) &\to \left(\alpha N - 1 - \frac{N-1}{2}\right)\log\left(\frac{\alpha}{1-\alpha}\right) + \frac{N-1}{2}\log\left(4\alpha(1-\alpha)\right) \\ &= (\alpha N - 1)\log\alpha + N(1-\alpha)\log(1-\alpha) + (N-1)\log 2.\end{aligned}$$

Thus,

$$\begin{aligned}f_N(y) &\sim P(y)^{\alpha N-1}(1-P(y))^{N-\alpha N}\frac{Np(y)2^N}{\sqrt{2\pi(N-1)4\alpha(1-\alpha)}}e^{-(\alpha N-1)\log\alpha - N(1-\alpha)\log(1-\alpha)-(N-1)\log 2} \\ &\sim p(y)\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\left(\frac{P(y)}{\alpha}\right)^{\alpha N-1}\left(\frac{1-P(y)}{1-\alpha}\right)^{N(1-\alpha)}.\end{aligned}$$

1.3 A "Central Limit Theorem" for Order Statistics

Let y_α be the mean of the random variable X_i and let α be chosen such that $P(y_\alpha) = \alpha$. Then a sort of "CLT" holds for the PDF of the order statistic $Y_N^{(\alpha)}$.

For large N, most of the probability density comes from the region around the maximum of $f_N^\alpha(y)$ because either $P(y)^{N\alpha}$ or $(1-P(y))^{N(1-\alpha)}$ will be very small away from the maximum. Thus, it makes sense to approximate $f_N^\alpha(y)$ by expanding around its maximum. Since, f_N^α decays to zero so rapidly around its maximum, it is appropriate to write it in exponential form and expand the exponential:

$$f_N^\alpha(y) = \sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}e^{NW(y)}$$

where

$$W(y) \equiv \left(\alpha - \frac{1}{N}\right)\log\left(\frac{P(y)}{\alpha}\right) + (1-\alpha)\log\left(\frac{1-P(y)}{1-\alpha}\right) + \frac{\log p(y)}{N}.$$

Since $f_N^\alpha(y)$ and $W(y)$ reach a maximum at the same value of y, we can find $y_{\max}$ using $W(y)$. Setting the derivative of $W(y)$ to zero, we have

$$0 = \left(\alpha - \frac{1}{N}\right)\frac{p(y_{\max})}{P(y_{\max})} - (1-\alpha)\frac{p(y_{\max})}{1-P(y_{\max})} + \frac{1}{N}\frac{p'(y_{\max})}{p(y_{\max})}.$$

In limit of large N, we can neglect the terms multiplied by $1/N$, so we find

$$p(y_{\max})\left(\alpha - P(y_{\max})\right) = 0$$

so that

$$P(y_{\max}) = \alpha$$

and

$$y_{\max} = y_\alpha.$$

Expanding $W(y)$ around $y_{\max}$ in the limit $N \to \infty$,

$$\begin{aligned} W(y) &\sim W(y_\alpha) + \frac{1}{2}\frac{d^2W}{dy^2}(y_\alpha)(y-y_\alpha)^2 + \dots \\ &\sim \frac{\log p(y_\alpha)}{N} - \frac{(p(y_\alpha))^2}{2\alpha(1-\alpha)}(y-y_\alpha)^2 + \dots \end{aligned}$$

Thus, to leading order,

$$f_N^\alpha(y) \sim \sqrt{\frac{N\,(p(y_\alpha))^2}{2\pi\alpha(1-\alpha)}} \exp\left(\frac{N\,(p(y_\alpha))^2}{2\alpha(1-\alpha)}(y-y_\alpha)^2\right)$$

which is the Gaussian that we would expect for a "CLT". Shifting and rescaling the PDF in the usual way, we find that the PDF for

$$z = \frac{Y_N^{(\alpha)} - y_\alpha}{\sigma_{Y_N}}$$

is a standard normal distribution with variance

$$\sigma_{Y_N}^2 = \frac{\alpha(1-\alpha)}{N\,(p(y_\alpha))^2}$$

Notice that the normalization automatically works out for this approximation.

1.4 Application: Comparison of Mean and Median Statistics

Suppose that X_i are IID standard normal random variables. Then $y_\alpha = 0$ and $\alpha = 1/2$. Therefore in the limit of large N, the order statistic $Y_N^{(0.5)}$ has a normal distribution with mean 0 and variance

$$\begin{aligned} \left(\sigma_{Y_N^{(0.5)}}\right)^2 &= \frac{1}{4N\,(p(0))^2} \\ &= \frac{\pi}{2N}. \end{aligned}$$

In contrast, the random variable

$$Z_N = \frac{1}{N}\sum_{i=1}^{N} X_i$$

has a normal distribution with variance,

$$(\sigma_{Z_N})^2 = 1/N.$$

Thus, we see that $\sigma_{Y_N^{(0.5)}} < \sigma_{Z_N}$. This result justifies the practice of using the sample mean, Z_N, instead the sample median, $Y_N^{(0.5)}$, when estimating the value of an unknown population mean. While both statistics provide good estimates of the population mean, Z_N is a better because it has a smaller variance.

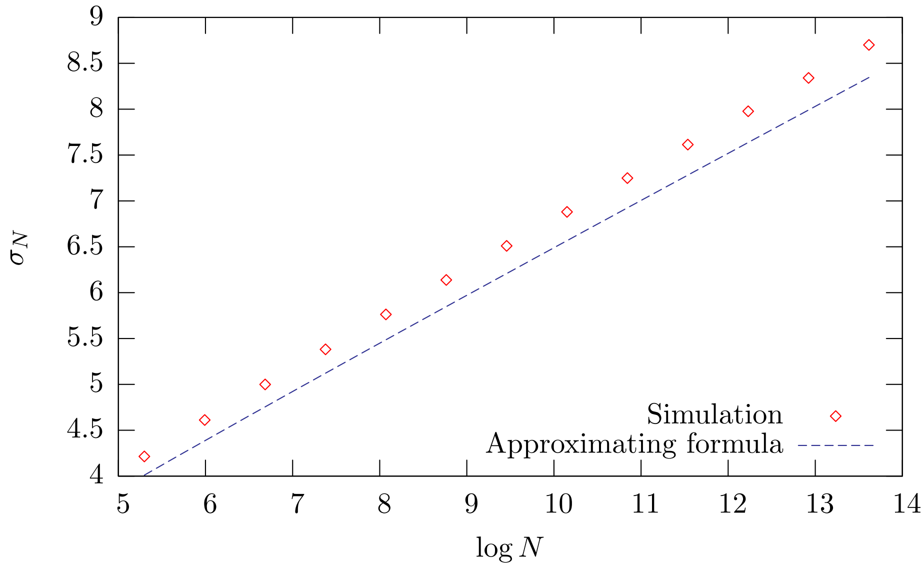

Figure 1: Comparison of the computed σ_N with the approximating formula. Values of N of the form 100×2^k for $k = 1, 2, \ldots, 13$ are shown in order for the data points to be equally spaced.

2 Winding angle of Pearson's walk

A C++ code to find the winding angle distribution of Pearson's random walk is shown in appendix A. This code was run for 10^7 random walks of length $N = 10^6$, and plot of σ is shown in figure 2. We see that the data points match the theoretical prediction

$$\left\langle \Theta_N^2 \right\rangle = \frac{1}{4}\log^2 N + \frac{4-\gamma}{2}\log N + O(1)$$

to a high degree of accuracy.

Figure 2 shows the computed PDF of winding distribution for $N = 10^6$. It is clear that this distribution has fairly large tails, and fitting a Gaussian

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

as shown in the diagram is not very satisfactory. A better choice would be the Cauchy distribution

$$g(x) = \lambda \frac{a}{\pi(a^2 + x^2)}.$$

Fitting over the range $|x| < 12$, we find that $\lambda = 1.27$ and $a = 6.76$, and we see from figure 2 that this matches extremely well in the central region. However there are still large deviations in the tails of the distribution, but since equation 2 tells us that σ_N scales logarithmically, this discrepancy may just be due to the fact that even $N = 10^6$ is not large enough to see a better match.

3 Globally Valid Asymptotics

Following the standard procedure for IID displacements, we write

$$\hat{p}(k) = \frac{1}{2a}\int_{-\infty}^{\infty} e^{-ikx-|x|/a} dx = \mathrm{Re}\int_0^{\infty} e^{-t-i(ka)t} dt = \frac{1}{1+(ka)^2},$$

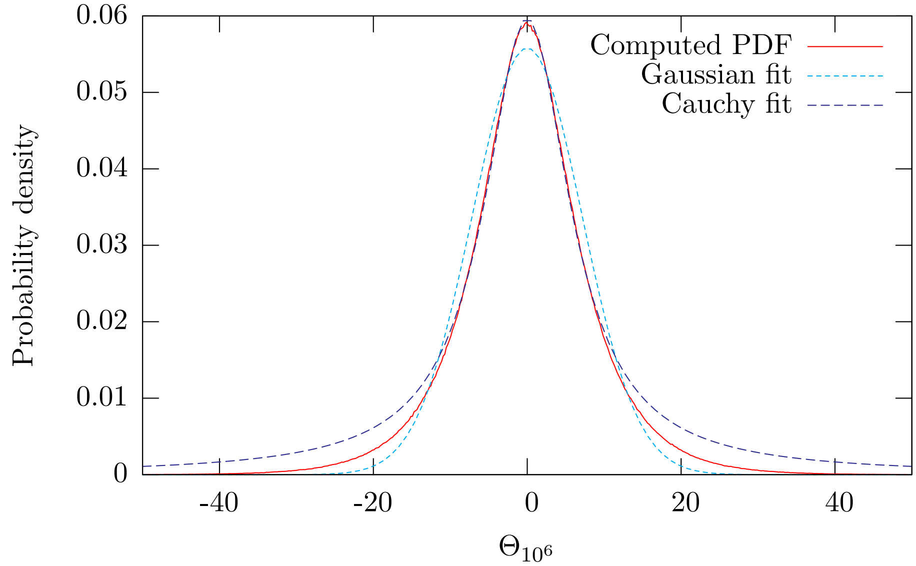

Figure 2: Computed PDF of Θ_{10^6} based on 10^7 random walks, with Gaussian and Cauchy fits.

so that

$$P_N(X) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{ikX}\left(\frac{1}{1+k^2a^2}\right)^N = \frac{1}{2\pi a}\int_{-\infty}^{\infty} dp\ e^{-N\left[-ip\zeta+\log\left(1+p^2\right)\right]},$$

where we define

$$\zeta = \frac{X}{Na}, \qquad p = ka.$$

Thus,

$$P_N\left(Na\zeta\right) = \frac{1}{2\pi a}\int_{-\infty}^{\infty} dp\ e^{-N\psi_N(p,\zeta)},$$

where $\psi_N(p,\zeta) = -ip\zeta + \log\left(1+p^2\right)$. Next, we explore the function $\psi_N(p,\zeta)$ in the complex p-plane for saddle-point(s):

$$\frac{\partial\psi}{\partial p} = 0 = -i\zeta + \frac{2p}{1+p^2} \qquad \Rightarrow \qquad p_s^{\pm} = \frac{-i}{\zeta}\left(1 \pm \sqrt{1+\zeta^2}\right).$$

The second derivative

$$\frac{\partial^2\psi}{\partial p^2} = 2\frac{1-p^2}{\left(1+p^2\right)^2}$$

is nonzero and is finite as long as $p \neq \pm i$. To evaluate the integral, we deform the contour as indicated in Fig. 1. In either case ($\zeta > 0$ and $\zeta < 0$) we pick only one of the saddle-points, namely p_-. The overall value of the integral is zero, since $e^{-N\psi_N(p,\zeta)}$ is analytic inside the contour. The integral over the (infinetely far) segments vanishes, so the integral along the real axis is equal to (minus) the integral along the displaced axis, which is dominated by the saddle point.[1] The result of the saddle-point contribution is written most compactly if we parametrize using

$$\zeta = \sinh t \qquad \Rightarrow \qquad p_- = -i\tanh\left(t/2\right),$$

[1] Strictly speaking, along the line parallel to the real axis, $\mathrm{Im}(\psi) \neq$ const, i. e. the phase is not constant. However, it is changing slowly in the vicinity of the saddle point and its influence is minor.

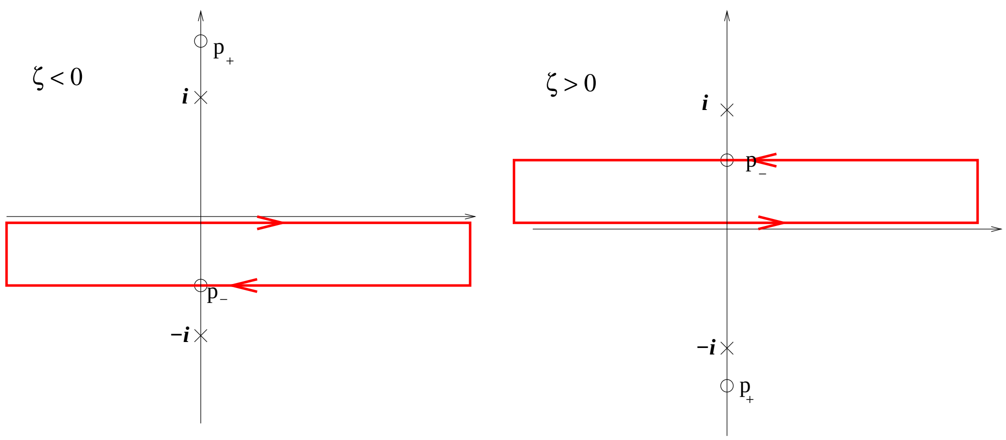

Figure 3: The saddle points and the integration contours.

so that

$$\psi_N\left(p_-\right) = \cosh t - 1 + \log\left(\cosh\left(t/2\right)\right)$$

and

$$\psi_N''\left(p_-\right) = \left(1 + \cosh t\right)\cosh t.$$

Plugging these equations into the expression for the saddle-point contribution, we obtain

$$P_N\left(t\right) = \frac{1}{2\pi a}\sqrt{\frac{2\pi}{N\left(1+\cosh t\right)\cosh t}}\exp\left[-N\left(\cosh t - 1 + \log\left[\cosh\left(t/2\right)\right]\right)\right].$$

If $\zeta \to 0$ then so does t, which leads to

$$P_N\left(X\right) \to \frac{1}{\sqrt{2\pi N(2a^2)}}\exp\left(-\frac{X^2}{2(2a^2)N}\right),$$

which is just the Central Limit Theorem for a distribution with zero mean and variance of $2a^2$. When $X \to \infty$,

$$P_N\left(X\right) \to e^{-N|\zeta|} = e^{-|X|/a}.$$

4 The Void Model

4.1 Equation for the void probability

Since a void at (M, N) has an equal probability of moving to $(M-1, N+1)$ or $(M+1, N+1)$, we know that

$$V_{N+1}(M) = \frac{V_N(M-1) + V_N(M+1)}{2}.$$

Expressing this in terms of $\eta(x, y) = \eta(Md, Nh) = V_N(M)$ gives

$$\eta(x, y+h) = \frac{\eta(x-d, y) + \eta(x+d, y)}{2}. \quad (2)$$

For h and d small, we can write

$$\eta + h\eta_y + \frac{h^2}{2}\eta_{yy} + \ldots = \frac{\eta + d\eta_x + \frac{d^2}{2}\eta_{xx} + \eta - d\eta_x + \frac{d^2}{2}\eta_{xx} + \ldots}{2}$$

which after simplification becomes

$$h\eta_y + \frac{h^2}{2}\eta_{yy} + \ldots = \frac{d^2}{2}\eta_{xx} + \ldots.$$

We see that in order to get a sensible answer in the limit of $h, d \to 0$, we must have $h \propto d^2$, and we therefore see it is appropriate to define a diffusion length of the form $b = d^2/2h$. Thus, as $h, d \to 0$, we obtain

$$\eta_y = b\eta_{xx}$$

which is a simple diffusion equation, but with the time variable replaced by the height, y.

4.2 Equation for the particle probability

To derive the recursion relation, we begin by considering a particle initially localized at (M, N). The ratio of the probability of it moving down-left to $(M-1, N-1)$ as opposed to down-right to $(M+1, N-1)$ is the same as the ratio of $V_{N-1}(M-1)$ to $V_{N-1}(M+1)$, and hence we see that

$$\begin{aligned} \mathbb{P}(\text{down-left}) &= \frac{V_{N-1}(M-1)}{V_{N-1}(M+1) + V_{N-1}(M-1)} = \frac{V_{N-1}(M-1)}{2V_N(M)} \\ \mathbb{P}(\text{down-right}) &= \frac{V_{N-1}(M+1)}{V_{N-1}(M+1) + V_{N-1}(M-1)} = \frac{V_{N-1}(M+1)}{2V_N(M)}. \end{aligned}$$

Thus, the probability of finding a particle at $(M, N-1)$ is just the probability of it being at $(M+1, N)$ times the probability of it moving down-left, plus the probability of it being at $(M-1, N)$ times the probability of it moving down-right:

$$P_{N-1}(M) = P_N(M-1)\frac{V_{N-1}(M)}{2V_N(M-1)} + P_N(M+1)\frac{V_{N-1}(M)}{2V_N(M+1)}. \quad (3)$$

4.3 Continuum limit of $P_N(M)$

If we define $\rho(x, y) = \rho(Md, Nh) = P_N(M)$, then equation 3 can be written in the form

$$\frac{2\rho(x, y-h)}{\eta(x, y-h)} = \frac{\rho(x-d, y)}{\eta(x-d, y)} + \frac{\rho(x+d, y)}{\eta(x+d, y)}.$$

Defining $\sigma = \rho/\eta$ we obtain

$$\sigma(x, y-h) = \frac{\sigma(x-d, y) + \sigma(x+d, y)}{2}$$

which is exactly the same as equation 2 but with η replaced by σ, and h replaced by $-h$. We therefore know immediately that σ satisfies a diffusion equation of the form

$$-\sigma_y = b\sigma_{xx}.$$

Substituting in for σ and expanding the partial derivatives produces

$$\frac{\eta_y \rho}{\eta^2} - \frac{\rho_y}{\eta} = b\left(\frac{\rho_{xx}}{\eta} - \frac{2\rho_x \eta_x}{\eta^2} + \frac{2\rho\eta_x^2}{\eta^3} - \frac{\rho\eta_{xx}}{\eta^2}\right).$$

By rearranging, and making use of the equation $\eta_y = b\eta_{xx}$, we obtain

$$\begin{aligned} \rho_y &= 2b\left(\frac{\rho_x\eta_x}{\eta} - \frac{\rho\eta_x^2}{\eta^2} + \frac{\rho\eta_{xx}}{\eta}\right) - b\rho_{xx} \\ &= 2b\frac{\partial}{\partial x}\left(\frac{\rho\eta_x}{\eta}\right) - b\rho_{xx}. \end{aligned}$$

4.4 Solution of the continuum equations

Taking the Fourier transform of the equation $\eta_y = b\eta_{xx}$ in the x variable gives

$$\begin{aligned} \tilde{\eta}_y(k,y) &= b(ik)^2\tilde{\eta}(k,y) \\ &= -bk^2\tilde{\eta}(k,y) \end{aligned}$$

and thus we obtain

$$\tilde{\eta}(k,y) = F(y)e^{-bk^2 y}$$

for some arbitrary function $F(y)$. The boundary condition $\eta(x,0) = \delta(x)$ tells us that $\tilde{\eta}(k,0) = 1$ and therefore we get

$$\tilde{\eta}(k,y) = e^{-bk^2 y}.$$

Taking the inverse Fourier transform, we obtain

$$\begin{aligned} \eta(x,y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{\eta}(k,y)\,dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-bk^2y+ikx}dk \\ &= \frac{e^{-\frac{x^2}{4by}}}{2\pi}\int_{-\infty}^{\infty} e^{-by\left(k-\frac{ix}{2by}\right)^2}dk \\ &= \frac{e^{-\frac{x^2}{4by}}}{\sqrt{4\pi by}}. \end{aligned}$$

From this result, we see that

$$\frac{\eta_x}{\eta} = -\frac{x}{2by}$$

and therefore the equation for ρ becomes

$$y\rho_y = -by\rho_{xx} - \frac{\partial}{\partial x}(\rho x).$$

To solve this equation, we first take the Fourier transform with respect to x, which yields

$$y\tilde{\rho}_y = -by(ik)^2\tilde{\rho} - (ik)\left(i\frac{\partial}{\partial k}\right)\tilde{\rho},$$

$$y\tilde{\rho}_y - k\tilde{\rho}_k = byk^2\tilde{\rho}.$$

This is a first order partial differential equation, so we can apply the method of characteristics. The characteristics are given by

$$\frac{dy}{y} = -\frac{dk}{k} = \frac{d\tilde{\rho}}{\tilde{\rho}byk^2}$$

from which we obtain

$$ky = C$$

and

$$\frac{\partial\tilde{\rho}}{\partial k} = -b\tilde{\rho}ky = -b\tilde{\rho}C \qquad \Longrightarrow \qquad \tilde{\rho} = De^{-k^2yb}$$

for some constants C and D. Thus the general solution is

$$\tilde{\rho}(k, y) = F(ky)e^{-k^2yb}.$$

To find F, we consider a boundary condition of the form $\rho(x, y_0) = \delta(x - x_0)$, which corresponds to a point particle initially localized at (x_0, y_0). Taking the Fourier transform of the boundary condition gives $\tilde{\rho}(k, y_0) = e^{-ikx_0}$ and hence

$$\begin{aligned} F(ky_0)e^{-k^2y_0b} &= e^{-ix_0k} \\ F(ky_0) &= e^{-ix_0k+k^2y_0b} \\ F(\lambda) &= \exp\left(\frac{\lambda^2b - ix_0\lambda}{y_0}\right). \end{aligned}$$

Thus

$$\tilde{\rho}(k, y) = \exp\left(-bk^2y\left(1 - \frac{y}{y_0}\right) - i\frac{x_0ky}{y_0}\right)$$

and taking the inverse Fourier transform gives

$$\begin{aligned} \rho(x, y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \exp\left(-bk^2\left(1 - \frac{y}{y_0}\right) + ik\left(x - \frac{x_0y}{y_0}\right)\right) \\ &= \frac{1}{\sqrt{4\pi by\left(1 - \frac{y}{y_0}\right)}} \exp\frac{-\left(x - \frac{x_0y}{y_0}\right)^2}{4by\left(1 - \frac{y}{y_0}\right)}. \end{aligned} \tag{4}$$

Thus we see that the probability density function of the particle position is a gaussian for all values of y, with mean and variance

$$\mu = \frac{x_0y}{y_0}, \qquad \sigma^2 = 2by\left(1 - \frac{y}{y_0}\right).$$

We see that as y decreases, the mean particle position moves linearly towards the orifice. The variance of the distribution starts at 0 when $y = y_0$ corresponding to the delta function initial condition; it then increases, reaching a maximum at $y = y_0/2$ before decreasing to 0 as y approaches 0, corresponding to the fact that the particle must exit through the point orifice at $x = y = 0$.

This answer also demonstrates that in the Void Model particles will diffuse at the same rate as the voids, since the same diffusion parameter b appears in both solutions. From equation 4, we see when the particle falls to half its original height, its probability density function has a comparable width to the velocity profile. In reality, this is extremely unphysical, since experiments show that particles tend to diffuse on a much smaller scale than the width of the velocity profile.

A Appendix: C++ code for question 2

```
#include <cstdio>
#include <iostream>
#include <fstream>
#include <cmath>
using namespace std;

const double pi=3.1415926535897932384626433832795,p=2*pi;
const long m=10000;      //Number of measurements
const long n=100;        //Number of steps per measurement
const long w=400000;     //Number of walkers

inline double arg(double x,double y) {
        return x+y>0?(x>y?atan(y/x):pi/2-atan(x/y)):
                        (x>y?-atan(x/y)-pi/2:atan(y/x)+(y>0?pi:-pi));
}

int main () {
        long i,j,k,q[4000];int g;
        double t,x,y,yy,s[m];
        for(k=0;k<m;k++) s[k]=0;
        for(i=0;i<4000;i++) q[i]=0;
        fstream file;
        for(j=0;j<w;j++) {
                x=1;y=0;i=1;g=0;
                for(k=0;k<m;k++) {
                        while(i++<n) {
                                yy=y+sin(t=double(rand())/RAND_MAX*p);
                                if(yy>0) {
                                        if(y>0) x+=cos(t);else
                                        if(x*yy-(x+=cos(t))*y<0) g--;
                                } else {
                                        if(y<=0) x+=cos(t);else
                                        if(x*yy-(x+=cos(t))*y>0) g++;
                                }
                                y=yy;
                        }
                        s[k]+=(t=arg(x,y)+g*p)*t;
                        i=0;
                }
                i=int(2000+t*10);if (i<0) i=0;if (i>=4000) i=3999;
                q[i]++;
        }
        file.open("berger",fstream::out|fstream::trunc);
        for(k=0;k<m;k++) {
```

```
                file << k+1 << "␣" << log(100.0*(k+1)) << "␣"
                     << s[k]/w << "␣" << sqrt(s[k]/w) << endl;
        }
        file.close();
        file.open("pdf",fstream::out|fstream::trunc);
        for(i=0;i<4000;i++) {
                file << (i-1999.5)/10 << "␣" << q[i] << endl;
        }
        file.close();
}
```