Rank probabilities for normal or gamma variates

Michael A. Newton[1]

December 2015

UW Madison Statistics Department Technical Report #??

[1]Departments of Statistics and of Biostatistics and Medical Informatics (Dept of Statistics; 1300 University Avenue; Madison, WI 53706; newton@stat.wisc.edu), University of Wiscsonsin-Madison, Madison, Wisconsin.

Abstract

The problem addressed is how to efficiently calculate the probability that independent but differently distributed normal or gamma random variables appear in a particular rank order. The computation arises in Bayesian inference when the underlying parameters are subject to ordering constraints. We show a case from the analysis of preference ordering and another case **clustering**. Constraints may also appear in the sample space, and then the computation may be required to evaluate the sampling probability of some event, as for example occurs in the evaluation of rank likelihood **or **. We confirm that the normal-rank probability is the limit of a series of gamma-rank probabilities. Using a compact formula for gamma-rank probabilities, we also show how a dynamic programming algorithm enables efficient numerical evaluation. **beyond ranks to up/down constraints**

KEYWORDS: gamma ranking; Viterbi algorithm .

1 Introduction

Under consideration is a finite collection of mutually independent real-valued random variables. In one case, they are normally distributed with possibly different means and variances. In another case, they are gamma distributed, with possibly different shapes and rates. The parameters are considered known in either case, and of interest is the probability that a particular ranking of the variables occurs. It is routine to evaluate this probability by Monte Carlo simulation; however a numerical solution is preferable in applications either if the probability itself is very small, or if very many instances of the probability need to be computed. This paper shows how the normal case can be approximated by the gamma case and also how a version of the Viterbi algorithm solves the gamma case.

Rank probabilities appear in various domains of statistics.

Preference ordering: A toy example from Bayesian inference is illustrative. A

population of voters is sampled and each voter reports which of three candidates is his/her favorite for an upcoming election. The population is partitioned into three subsets, with unknown proportions (p, q, r) of voters favoring each of the three candidates. Similarly the sample is partitioned into three observed counts, (a, b, c), say. The likelihood $L(p, q, r)$ from simple random sampling is $L(p, q, r) = p^a q^b r^c$. An analyst may be interested in preference orderings, such as $E = (p > q > r)$, or one of the five other versions, as being relevant properties of the population under study. The posterior probability of E involves integrating a normalized version of L times a prior density over the event E. Taking a flat prior, for example, this posterior density is simply the density of a Dirichlet distribution with shapes $(a + 1, b + 1, c + 1)$. It is well known that the Dirichlet vector (p, q, r) is equal in distribution to a vector $(V_1/S, V_2/S, V_3/S)$, where $S = V_1 + V_2 + V_3$, and where the V_i's are independent gamma-distributed variables with common rate and shapes $(a + 1, b + 1, c + 1)$. Thus to compute the posterior probability of a preference ordering in the population, it is equivalent to compute the probability that a sequence of independent gamma variates goes in descending order. For instance, if $n = 30$ voters gave results $(a, b, c) = (12, 10, 8)$, it turns out that the posterior probability is 0.38 that the population matches this empirical ordering.

less toy from Tom Louis??

Seration:

Other:

Rank likelihood: Consider a two-sample problem involving independent measure-

ments from two source populations, with cumulative distribution functions F and G. Inferring differences between F and G is one of the classical problems in statistics. An interesting semiparametric method for this problem is based on rank likelihood (Pettitt, 1982; Doksum, 1987; Dabrowska, Doksum, and Miura, 1988; Cuzick, 1988; Tsukahara, 1997). Ranks provide a natural and robust way to test the null hypothesis that $F = G$ (e.g., Hoeffding, 1951; Kuk, 2009), but, in part for computational reasons, they have been less well developed beyond the null. Rank-likelihood is the marginal probability of the ranks, expressed as a function of parameters. It has been incredibly important for right-censored survival data (in the Cox proportional hazards model), and may have further potential for other data structures.

Rank likelihood requires a semiparametric model for data. Various schemes are possible, but a compelling one is to assume that some unknown monotone transformation makes the data normally distributed with constant variance and with a population-specific mean. The rank likelihood eliminates the unknown, nuisance transformation, and involves a single parameter measuring the difference between means on the transformed scale. To evaluate the rank likelihood is to evaluate the probability that transformed data (*i.e.*, normal variates) achieve a particular ordering. A rank likelihood for a similar gamma-based semiparametric model provides an approximation to the normal case (Section **).

Mixture-based clustering:

2 Gamma approximation to normal

When its shape parameter is large, the gamma distribution is approximately normal. This property can be used to advantage to approximate normal rank probabilities.

Suppose we know real-valued means $\mu_1, \mu_2, \ldots, \mu_K$ and positive variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$, and $Z_1, Z_2, \ldots, Z_K$ are mutually independent normal variables with $Z_k \sim \text{Normal}(\mu_k, \sigma_k^2)$. Of interest is the probability of a given rank ordering. Without loss of generality, consider

$$P\left(E_{\text{normal}}\right) = P(Z_1 > Z_2 > \ldots > Z_K). \tag{1}$$

Of course, $P\left(E_{\text{normal}}\right) = 1/K!$ if all the means are equal and all variances are equal. When $K = 2$, the probability can be written as a tail probability for one normal arising from the difference $Z_1 - Z_2$, and so evaluation is also elementary in this case. In general, however, the probability depends in a complicated way on the means and variances. **citations**

A series of probabilities from gamma-distributed variables can be used to approximate $P\left(E_{\text{normal}}\right)$. Recall that a gamma distribution with shape $\alpha > 0$ and rate $\lambda > 0$ has probability density proportional to $v^{\alpha-1}\exp\{-v\lambda\}$ for positive real v. For each $n = 1, 2, \ldots$, consider mutually independent variables $V_{n,1}, V_{n,2}, \ldots, V_{n,K}$ where $V_{n,k}$ is gamma distributed with shape parameter $\alpha_{n,k} = n/\sigma_k^2$ and rate parameter $\lambda_{n,k} = \left(\frac{n}{\sigma_k^2}\right)\left\{1 - \frac{\mu_k}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right\}$. In comparison to (1), the gamma-rank probability is

$$P\left(E_{n,\text{gamma}}\right) = P\left(V_{n,1} > V_{n,2} > \ldots > V_{n,K}\right). \tag{2}$$

Theorem 1. *With the shape and rate parameters as defined above,*

$$\lim_{n\to\infty} P\left(E_{n,\text{gamma}}\right) = P\left(E_{\text{normal}}\right).$$

Thus a scheme to compute gamma-rank probabilities can be used to approximate the normal case.

In what follows we suppress the approximation level n from the notation, and write things for gamma variables $V_1, V_2, \ldots, V_K$ with shapes $\alpha_1, \alpha_2, \ldots, \alpha_K$ and rates $\lambda_1, \lambda_2, \ldots, \lambda_K$. When $K = 2$, the probability (2) is equivalent to one for a single beta-distributed variable, and so the computation of $P\left(E_{\text{gamma}}\right)$ is standard. It seems that the general case has not received much attention. Recently, Newton and Chung (2009) reported the following formula for $P\left(E_{\text{gamma}}\right)$ in the case of integer-valued shape parameters:

$$P\left(E_{\text{gamma}}\right) = \sum_{m_1=0}^{\alpha_1-1} \sum_{m_2=0}^{m_1+\alpha_2-1} \cdots \sum_{m_{K-1}=0}^{m_{K-2}+\alpha_{K-1}-1} p_1(m_1)\, p_2(m_2) \cdots p_{K-1}(m_{K-1}). \quad (3)$$

where the probability mass function $p_k(\cdot)$ is that of a negative binomial random variable M_k with shape α_{k+1} and scale $(\lambda_1+\lambda_2+\cdots+\lambda_k)/\lambda_{k+1}$, which corresponds to the mass function:

$$p_k(m) = \frac{\Gamma(m+\alpha_{k+1})}{\Gamma(\alpha_{k+1})\,\Gamma(m+1)} \left(\frac{\lambda_{k+1}}{\Lambda_{k+1}}\right)^{\alpha_{k+1}} \left(\frac{\Lambda_k}{\Lambda_{k+1}}\right)^{m} \quad (4)$$

for integers $m \geq 0$. Here $\Lambda_k = \sum_{j=1}^{k} \lambda_j$. Representation (3) was found by embedding the gamma variables in a certain collection of Poisson processes and then seeing the ordering event as equivalent to an event on related count variables.

It may happen that the first $K-1$ shape parameters α_k equal one, in which case the sum in (3) collapses to a single term. Beyond this special exponential case an

explicit summation has to be evaluated. Newton and Chung (2009) showed how the sum-product algorithm could be used to redistribute elements of (3) in order to evaluate $P(E)$ relatively efficiently (as in Kschischang *et al.* 2001) . The sum-product formulation speeds up the naive summation, especially as K increases, just as the Baum-Welch forward-backward algorthm speeds up hidden-Markov computatons.

When the summands in (3) are very small, there is numerical instability in the evaluation of $P(E)$, even when using the sum-product algorithm. Here we show how the Viterbi algorithm can be applied to identify the modal summand. By factoring out the modal summand one gets an improved evaluation of $\log P(E)$. We also present an R package `grankp` for implementing the computation.

3 The Viterbi connection

Although there is no hidden Markov model (HMM) in the system of gamma variables, the summation (3) is structurally equivalent to that of a marginal probability for an HMM. Obviously there are finitely many terms in the sum, even though the negative binomial mass functions are defined for all natural numbers, and it is convenient to define new functions

$$\begin{aligned} h_1(m_1) &= p_1(m_1)\, 1[m_1 \leq \alpha_1 - 1] \\ h_2(m_1, m_2) &= p_2(m_2)\, 1[m_2 \leq m_1 + \alpha_2 - 1] \\ &\vdots \\ h_{K-1}(m_{K-2}, m_{K-1}) &= p_{K-1}(m_{K-1})\, 1[m_{K-1} \leq m_{K-2} + \alpha_{K-1} - 1]. \end{aligned}$$

Since no index m_k can exceed $A = \alpha_1 + \alpha_2 + \ldots + \alpha_{K-1} - (K-1)$, it is reasonable to think of each of the $K-1$ sums in (3) as restricted to $0, 1, \ldots, A$. Re-writing (3),

$$P(E) = \sum_{m_1=0}^{A} \sum_{m_2=0}^{A} \cdots \sum_{m_{K-1}=0}^{A} h_1(m_1) \prod_{k=2}^{K-1} h_k(m_{k-1}, m_k). \tag{5}$$

The linkages $h_k(m_{k-1}, m_k)$ between neighboring indices give the summation the same character as the summation required to obtain a marginal probability in an HMM.

Recall the *log-sum-exp* formula for evaluating the logarithm of a summation when the summands are subject to numerical underflow (e.g., Press *et al.* 2007, page 844):

$$\log\left(\sum_i \exp(u_i)\right) = \max_j(u_j) + \log\left(\sum_i \exp[u_i - \max_j(u_j)]\right).$$

To apply this formula in the present context, we must obtain the vector $(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_{K-1})$ in the integer lattice $[0, A]^{K-1}$ that maximizes the summand in (5). We develop the Viterbi algorithm for this purpose as follows (*c.f.* Rabiner 1989). For any $k \geq 2$ and $k \leq K-1$, we define the multi-variate function

$$T_k(m_1, m_2, \ldots, m_k) = \log h_1(m_1) + \sum_{j=2}^{k} \log h_j(m_{j-1}, m_j),$$

where the h_j's, from above, are truncated probability mass functions, and allowing $\log(0) = -\infty$. These are sub-components of the summands in (5), presented on the logarithmic scale. Note that the final function $T_{K-1}(m_1, \ldots, m_{K-1})$ equals the log of the entire summand, and so the optimizer of T_{K-1} equals the required mode $(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_{K-1})$. From the multi-variate T_k's we derive a sequence of

univariate functions (i.e., vectors) $\delta_2, \delta_3, \ldots, \delta_{K-1}$ as

$$\begin{aligned}
\delta_2(m) &= \max_{m_1} T_2(m_1, m) \\
\delta_3(m) &= \max_{m_1, m_2} T_3(m_1, m_2, m) \\
&\vdots \\
\delta_k(m) &= \max_{m_1, m_2, \ldots, m_{k-1}} T_k(m_1, m_2, \ldots, m_{k-1}, m) \\
&\vdots
\end{aligned}$$

where, again, each m, m_k takes values in $0, 1, \ldots, A$. On the surface it seems that the computation of each $\delta_k(m)$ requires a search over $(A+1)^{k-1}$ values for T_k, but owing to the pairwise linkages the computation can be done recursively much more quickly than that. One fact driving the recursion is that

$$T_k(m_1, \ldots, m_k) = T_{k-1}(m_1, \ldots, m_{k-1}) + \log h_k(m_{k-1}, m_k).$$

Additionally, maximization over multiple arguments can be done in two steps, by maximizing in one set of arguments for each value of the other, and then by maximizing this reduced profile in the other arguments. Put together,

$$\delta_k(m) = \max_{m_{k-1}} \left\{ \max_{m_1, \ldots, m_{k-2} | m_{k-1}} \left[T_{k-1}(m_1, \ldots, m_{k-1}) + \log h_k(m_{k-1}, m) \right] \right\}.$$

The inner optimization is over $k-2$ variables and keeps m_{k-1} fixed. Thus the second term inside the optimization is constant, and can be moved out:

$$\begin{aligned}
\delta_k(m) &= \max_{m_{k-1}} \left\{ \log h_k(m_{k-1}, m) + \max_{m_1, \ldots, m_{k-2} | m_{k-1}} T_{k-1}(m_1, \ldots, m_{k-1}) \right\} \\
&= \max_{m_{k-1}} \left\{ \log h_k(m_{k-1}, m) + \delta_{k-1}(m_{k-1}) \right\}.
\end{aligned}$$

This shows that the sequence of δ_k vectors can be computed recursively, with a univariate optimization at each step. During this recursion one keeps track of arguments where the maxima occur, storing them in vectors ψ_k, say:

$$\psi_k(m) = \arg\max_{m_{k-1}} \{\log h_k(m_{k-1}, m) + \delta_{k-1}(m_{k-1})\} .$$

Completing the Viterbi algorithm, backtracking is used to recover the mode:

$$\begin{aligned} \hat{m}_{K-1} &= \arg\max_m \delta_{K-1}(m) \\ \hat{m}_k &= \psi_{k+1}(\hat{m}_{k+1}) \qquad \text{for } k = K-2, K-3, \ldots, 1 . \end{aligned}$$

4 Examples

4.1 Dirichlet ordering

We illustrate the Viterbi computation using the Dirichlet ordering problem from the Section 1. Of interest is the posterior probability that the population preference ordering for three candidates matches the empirical ordering based on observed counts $(12, 10, 8)$. The counts gives shapes $(\alpha_1, \alpha_2, \alpha_3) = (13, 11, 9)$ and the Dirichlet structure gives $\lambda_k = 1$ for $k = 1, 2, 3$. Simplifying (3), the event probablity $P(E)$ is

$$P(E) = \sum_{m_1=0}^{12} \sum_{m_2=0}^{m_1+10} p_1(m_1)\, p_2(m_2)$$

where, for $m \geq 0$,

$$\begin{aligned} p_1(m) &= \frac{\Gamma(m+11)}{\Gamma(m+1)\Gamma(11)} \left(\frac{1}{2}\right)^{11} \left(\frac{1}{2}\right)^{m} \\ p_2(m) &= \frac{\Gamma(m+9)}{\Gamma(m+1)\Gamma(9)} \left(\frac{1}{3}\right)^{9} \left(\frac{2}{3}\right)^{m} \end{aligned}$$

Reformatting the computation calls for both sums to range from 0 to $A = 22$ and requires objects

$$\log h_1(m_1) = \begin{cases} -\infty & \text{if } m_1 > 12 \\ \log p_1(m_1) & \text{else} \end{cases}$$

and

$$\log h_2(m_1, m_2) = \begin{cases} -\infty & \text{if } m_2 > m_1 + 10 \\ \log p_2(m_2) & \text{else} \end{cases}$$

Continuing, the bivariate $T_2(m_1, m_2)$ is the sum of these two functions, and the vector δ_2 has entries $\delta_2(m) = \max_{m_1} T_2(m_1, m)$. One finds that $\hat{m}_2 = 15$ is the argument maximizing $\delta_2(m)$, and then, backtracking, that the argmax is $\hat{m}_1 = 9$ in the computation of $\delta_2(15)$. It happens in this case that $\hat{m}_1 = 9$ and $\hat{m}_2 = 15$ separately maximize $p_1(m_1)$ and $p_2(m_2)$; in general the shape constraints can prohibit the joint mode from occurring at the conjunction of marginal modes.

To utilize the *log-sum-exp* formula for computing $\log P(E)$, note

$$\log P(E) = \log p_1(\hat{m}_1) + \log p_2(\hat{m}_2) + \log \left\{ \sum_{m_1=0}^{12} \sum_{m_2=0}^{m_1+10} \frac{p_1(m_1)\, p_2(m_2)}{p_1(\hat{m}_1)\, p_2(\hat{m}_2)} \right\}$$

4.2 Ordering genetic markers

4.3 Rank likelihood

Consider a two-sample problem with independent measurements $X_1, X_2, \ldots, X_n$. Let $\delta_1, \delta_2, \cdots, \delta_n$ denote known binary indicators that distinguish the two source populations, and denote the distribution of X_i by F when $\delta_i = 0$ and by G when $\delta_i = 1$. **start with Gamma ? ** The semiparametric model driving the calculation is to express each observation in terms of a normal variable. Let $Z_1, Z_2, \ldots, Z_n$ be independent and identically distributed from a normal distribution with shape a and rate a (thus they have mean 1 and variance $1/a$). (The approximate normality of this distribution, for large a, will be used to advantage in Section 4.) Assume that for some continuous monotone increasing function $h()$, and for some $\lambda_a > 0$,

$$X_i = h\left(\lambda_a^{\delta_i} Z_i\right). \tag{6}$$

The model is not so restrictive as it might seem, due to the absence of any parametric assumptions on h. Either F or G can take any continuous distribution; the substantive restriction enters in the allowable differences between the distributions. To appreciate model (6) further, consider $\theta = P(X_i < X_j)$ for i in one group and j in the other (so $\delta_i = 0$ and $\delta_j = 1$). Of course $\theta = 1/2$ when $F = G$; indeed θ is a

parameter we can hope to infer from the ranks alone. We have

$$\begin{aligned} \theta &= P(X_i < X_j) \\ &= P\left(h(Z_i) < h(\lambda_a Z_j)\right) \\ &= P\left(Z_i < \lambda_a Z_j\right) \\ &= P\left(B < \lambda_a(1-B)\right) \\ &= P\left(B < \frac{\lambda_a}{1+\lambda_a}\right) \end{aligned} \tag{7}$$

where B is a Beta-distributed variable with both shapes equal to a. Thus λ_a and θ are directly related.

Information about θ is contained in the ordering of the observations, say

$$E = \left\{X_{\pi_1} < X_{\pi_2} < \cdots < X_{\pi_n}\right\}.$$

The permutation $(\pi_1, \pi_2, \cdots, \pi_n)$ holds the so-called *anti-ranks* (they're the values returned by the `R` function `order`.) The rank likelihood for θ is

$$\begin{aligned} \mathrm{RL}(\theta) &= P(E) \\ &= P\left(X_{\pi_1} < X_{\pi_2} < \cdots < X_{\pi_n}\right) \\ &= P\left(\lambda_a^{\delta_{\pi_1}} Z_{\pi_1} < \lambda_a^{\delta_{\pi_2}} Z_{\pi_2} < \cdots < \lambda_a^{\delta_{\pi_n}} Z_{\pi_n}\right) \end{aligned} \tag{8}$$

owing to model (6). The $\lambda_a^{\delta_{\pi_j}}$ factors are constants in the calculation, and so $\mathrm{RL}(\theta)$ is expressible as the probability of an ordering of independent Gamma-distributed variables. The algorithm from ** obtains.

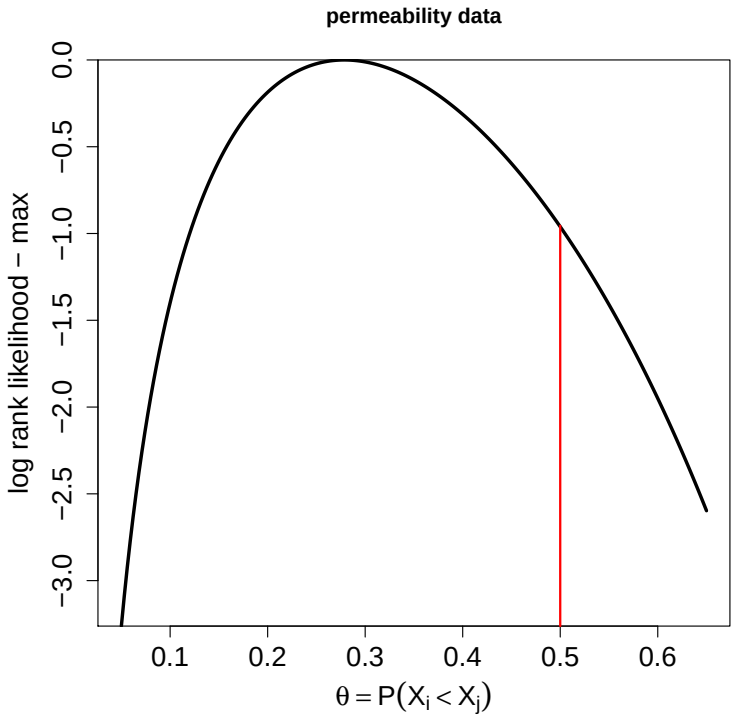

Figure* demonstrates rank likelihood for the permeability example from Hollander and Wolfe (1999), (example 4.1, page 110). A nominal 95% likelihood-based confidence interval for θ is $(0.08, 0.60)$ (dropping 2 log units from maximum), which differs from other approaches. The interval by Halperin, Gilbert, and Lachin (1987) is $(0.13, 0.55)$ and by Sen (1967) is $(0.02, 0.58)$, as reported in Hollander and Wolfe (1999) (page 131). (See also `?wilcox.test` in `R`.)

Numerical experiments showed that rank likelihood $\mathrm{RL}(\theta)$ is relatively insensitive to the shape parameter a used in (8). Since the Gamma distribution is approximately normal for large shapes, (3) may provide a useful numerical approach to long intractable normal rank integral (citations). The precise connection to nor-

mality is presented in Section 4. Roughly speaking, model (6) can be restated to say that some unknown monotone transformation converts our data into normally distributed scores. Further, taking these scores to be standard normal in one sample ($\delta_i = 0$), involves that in the other sample ($\delta_i = 1$) the scores are normal with mean $\sqrt{2}\Phi^{-1}(\theta)$ and unit variance. Here Φ is the standard normal cumulative distribution and θ is as in (7). (The phenomena is reminiscent of the bootstrap percentile method, which is justified when some unknown monotone transform of the parameter estimator has a constant-variance normal distribution...citation....)

**maybe equations and figure showing cdf G when $F(x) = x$ on $(0, 1)$. **

5 Discussion

... computation ...

...rank likelihood more generally

Acknowledgements

... R version 2.10.1

Proof of theorem 1

Each $V_{n,k}$ is approximately normal for large shape, regardless of constraints on the rate parameters.

Lemma: *Fix k, let $\alpha_{n,k} \to \infty$ as $n \to \infty$, and take any positive rates $\lambda_{n,k}$. As $n \to \infty$,*

$$W_{n,k} = \frac{V_{n,k} - \alpha_{n,k}/\lambda_{n,k}}{\sqrt{\alpha_{n,k}/\lambda_{n,k}^2}} \longrightarrow_d \text{Normal}(0,1).$$

This can be confirmed by showing that the characteristic function of $W_{n,k}$,

$$\begin{aligned} \phi_n(t) &= E\left\{\exp(itW_{n,k})\right\} \\ &= \left[\exp\left(-it\sqrt{\alpha_{n,k}}\right)\right]\left(1 - it/\sqrt{\alpha_{n,k}}\right)^{-\alpha_{n,k}}, \end{aligned}$$

converges to $\exp(-t^2/2)$.

To establish the theorem, consider the event $E_{j,k}^n = [V_{n,j} > V_{n,k}]$ for distinct j, k in $1, 2, \ldots, K$. Clearly, $E_{j,k}^n$ is equivalent to:

$$W_{n,j}\sqrt{\frac{\alpha_{n,j}}{\lambda_{n,j}^2}} + \frac{\alpha_{n,j}}{\lambda_{n,j}} > W_{n,k}\sqrt{\frac{\alpha_{n,k}}{\lambda_{n,k}^2}} + \frac{\alpha_{n,k}}{\lambda_{n,k}}. \tag{9}$$

The particular choices $\alpha_{n,k} = n/\sigma_k^2$ and $\lambda_{n,k} = \frac{n}{\sigma_k^2}\left\{1 - \frac{\mu_k}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right\}$ now come into play. Multiply the both sides of (9) by $\sqrt{n}$. While the W's are stablizing at standard normals (by lemma), the multiplicative factors are also converging. On the left,

$$\sqrt{n}\sqrt{\frac{\alpha_{n,j}}{\lambda_{n,j}^2}} \longrightarrow \sigma_j$$

as $n \to \infty$; and similarly on the right side of (9) the factor multiplying $W_{n,k}$ converges to σ_k. Some care is needed, since multiplication by $\sqrt{n}$ has also inflated the additive terms on both sides of (9). If we further subtract $\sqrt{n}$ from both sides, the

additive constant, on the left, say, is

$$\begin{aligned}\sqrt{n}\left(\frac{\alpha_{n,j}}{\lambda_{n,j}}\right) - \sqrt{n} &= \sqrt{n}\frac{n}{\sigma_j^2}\left[\frac{n}{\sigma_j^2}\left\{1 - \frac{\mu_j}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right\}\right]^{-1} - \sqrt{n} \\ &= \mu_j + o(1).\end{aligned}$$

In other words,

$$P(E_{j,k}^n) = P\left\{Z_{n,j} + o(1) > Z_{n,k} + o(1)\right\}$$

where $Z_{n,j} \longrightarrow_d$ Normal(μ_j, σ_j^2), $Z_{n,k} \longrightarrow_d$ Normal(μ_k, σ_k^2), and the $o(1)$ terms represent deterministic sequences converging to zero, all as $n \to \infty$. Finally, put the pairwise events together, noting $E_{n,\text{gamma}} = \bigcap_{k=2}^{K} E_{k-1,k}^n$, to establish the result. □