Dedictated to Jerry Lieberman on his 70th birthday.

An exponential inequality for U-statistics with applications to testing *

Peter J. Bickel Ya'acov Ritov

March 21, 2001

Abstract

We present a new exponential inequality for degenerate U-statistics. The bound of the log of the hazard is quadratic for small to medium values of the deviation and linear for larger value. We apply this bound to a family of test statistics and provide the key step in a optimality result for adaptive tests (Bickel and Ritov, 1992).

*Research partially supported by NSA Grant MDA 904-94-H2020 and the US/Israel Bi-National Science Foundation Grant 90-00031/2

1 Introduction and statement of the main result

Let $h(\cdot\,,\cdot\,)$ be a kernel such that $h(x,y) = h(y,x)$ for all x and y. Let $X, X_1, \dots, X_n$ be iid $U(0,1)$ random variables. We assume that the kernel satisfies the following conditions

$$\mathrm{E}\, h(\cdot\,, X) = 0, \qquad \|h\|_\infty = b,$$

for some $b < \infty$, where $\|h\|_\infty = \sum_{x,y} |h(x,y)|$. We prove here an exponential bound on the deviations of the U-statistics

$$U_n = n^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} h(X_i, X_j).$$

It is well known (cf. Serfling (1980)) that the asymptotic distribution of U_n is the same as the distribution of $\sum_{m=1}^{\infty} \gamma_m (Z_m^2 - 1)$ where $Z_1, Z_2, \dots$ are iid standard normal random variables and $\gamma_1, \gamma_2, \dots,$ are the eigenvalues (including multiplicities) of h considered as an operator $L_2[0,1] \to L_2[0,1]$ given by $hf(\cdot) = \int_0^1 h(\cdot\,, x) f(x)\, dx$. In particular, if $\gamma_m = k^{-1/2}$, $m = 1, 2, \dots, k$ and 0 otherwise, we obtain that the asymptotic distribution is, up to scale and location, χ_k^2. One could like to have a bound on the tail probabilities of U_n which is of the same order as the tail probabilities of the asymptotic distribution. In particular, one would like $-\log \mathrm{P}(U_n > y)$ to be quadratic for $y \le \sqrt{k}$ and linear for larger deviations. We will establish such bounds (Corollary 1) under a condition on the relative magnitude of h in two norms.

Let $\|g\|_* = \text{esssup}_x (\int_0^1 g^2(x,y)\,dy)^{1/2}$. Since $h(\cdot\,,\cdot\,)$ is bounded and symmetric it has a spectral decomposition,

$$h(x,y) = \|h\|_* \sum_{i=1}^{\infty} \nu_m \phi_m(x)\phi_m(y),$$

where ϕ_m, ν_m, $m = 1, 2, \ldots$ are all real. Since $\int_0^1 \phi_i(x)\phi_j(x)\,dx = \delta_{ij}$, we obtain that $\sum_{i=1}^{\infty} \nu_m^2 = \|h\|_2^2 / \|h\|_*^2 \le 1$. Let $\rho(h) = \max_m |\nu_m|$.

Theorem 1.1 *Define α_ε by $\alpha_\varepsilon \exp(\alpha_\varepsilon) = 3\varepsilon$. For any y, β, and d_n such that $y > 0$, $\rho^{-1} \ge \beta > 0$, and $\alpha_\varepsilon \sqrt{n}(e^{-\beta\rho}\|h\|_\infty/\|h\|_* + \beta)^{-1} > d_n > 0$. Then*

$$\begin{aligned}
&\mathrm{P}\left(\frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1} h(X_i, X_j) > y\right) \\
&\le \quad \exp\left\{-\frac{\beta e^{-\beta\rho}}{\|h\|_*} y + \frac{1}{2}\beta^2 + \frac{1}{2} C_1 n e^{-1/4(1-\varepsilon)d_n^2} + \frac{1}{n^{1/2}}\left(\frac{\beta^2(1+\beta e)}{2n^{1/2}} + d_n\beta\right)^3\right\} \\
&\quad + 3n \exp\left\{-\frac{1}{4}(1-\varepsilon)d_n^2\right\},
\end{aligned}$$

where $C_1 = \beta e^{-\beta\rho}\|h\|_\infty/\|h\|_* + \beta^2$.

The next corollary gives a more useful result.

Corollary 1.1 *Suppose that $\|h\|_\infty/\|h\|_* < n^{1/2-\eta}$ for $0 < \eta < 1/14$ then for every $\xi > 0$, $c > 2(e/2)^3$, and $\zeta > 1$ there is n_0, n_0 depends only on η, ξ, c, and ζ:*

$\mathrm{P}(U_n > y)$

$$\leq \begin{cases} \zeta \exp\left\{-\frac{e^{-2}y^2}{2(1+c\xi^7)\|h\|_*^2}\right\} + a_n & \frac{e^{-1}y}{(1+c\xi^7)\|h\|_*} \leq \rho^{-1}(h) \bigwedge \xi n^{2/7} \\ \zeta \exp\left\{-\frac{e^{-1}y}{2\|h\|_*}\left(\frac{1}{\rho(h)} \bigwedge n^{2/7}\right)\right\} + a_n & \text{otherwise} \end{cases}.$$

for every y and $n > n_0$ where $a_n = 3n\exp\{-\frac{1}{4}(1-\varepsilon)n^{2\eta}\}$.

Proof Take $d_n = n^\eta$,

$$\beta = \min\{\rho^{-1}, \frac{e^{-1}y}{(1+c\xi^7)\|h\|_*}, \xi n^{2/7}\}.$$

and note that for $\beta < \xi n^{2/7}$

$$\begin{aligned} \frac{1}{n^{1/2}}\left(\frac{\beta^2(1+\beta e)}{2n^{1/2}} + d_n\beta\right)^3 &\leq \beta^2\left(\frac{e}{2}\xi^{7/3} + \frac{\xi^{4/3}}{n^{2/7}} + \frac{\xi^{1/3}}{n^{1/14-\eta}}\right)^3 \\ &\leq \tfrac{1}{2}c\xi^7\beta^2, \qquad n > n_0 \end{aligned}$$

for n_0 large enough.

□

A weaker bound for weaker conditions is given by the next corollary.

Corollary 1.2 *Suppose that for some $\eta > 0$: $y/\|h\|_* \leq \eta n^{1/6}/\log(n)$, $y/\|h\|_* \leq 1/\rho(h)$, and $\|h\|_\infty/\|h\|_* < \eta\sqrt{n/\log(n)}$. Then for all $\gamma > 0$ there are n_0, and ξ which are functions of η and γ only such that for all $n > n_0$*

$$\mathrm{P}\left(U_n > y\right) \leq (1+\xi)\exp\left\{-\lambda\left(y/\|h\|_*\right)^2\right\} + \xi n^{-\gamma}.$$

Proof Take again $d_n = c_1 \log(n)$ and $\beta = y/\|h\|_*$.

$\square$

Many empirical process type of results for the U-statistics appeared recently beginning with Nolan and Pollard (1987, 1988). De La Pena (1992) proved an important decoupling and Khintchines inequality. A large deviation principle for U-statistics was proved by Eichelsbacher and Löwe (1993). Our result appears to give a different information.

The proof of the theorem is given in the next section. The application to testing is given in the third section.

2 Proof of Theorem 1

Let $\mathcal{F}_i = \sigma(X_1, X_2, \ldots, X_i)$ and $\tilde{W} = \sum_{j=1}^{i-1} h(X_i, X_j)$, $i = 1, \ldots, n$. Note that $\mathrm{E}(\tilde{W}_i \,|\, \mathcal{F}_{i-1}) = 0$ and hence $U_i = \sum_{j=2}^{i} \tilde{W}_i$ is a martingale with respect to the filtration $\{\mathcal{F}_i\}$. The $\tilde{W}_i$'s themselves, being a sum of bounded iid random variables can easily be bounded. So, it is possible to use methods useful for bounding the sum of martingale differences sequences. We give its proof since the main result uses an extension of the same idea.

Lemma 2.1

i. *For any random variable* X *such that* $\mathrm{P}(|X|) > b) = 0$,

$$\mathrm{E}\, e^X \le \exp\left\{ \mathrm{E}\, X + \frac{1}{2} \operatorname{Var}(X) + \frac{1}{6} \operatorname{Var}(X) e^b \right\}.$$

ii. *Let* $Y_1, Y_2, \ldots, Y_n$ *be a martingale difference sequence and let* $\mathcal{F}_i$ *be the minimal* σ*-field such that* $Y_1, Y_2, \ldots, Y_i$ *are measurable* $\mathcal{F}_i$*. Assume that* $\operatorname{Var}(Y_i \,|\, \mathcal{F}_{i-1}) = v_i$ *(* v_i *non-random),* $\mathrm{P}(|Y_I| \le b \,|\, \mathcal{F}_{i-1} = 1$ *for all* $i = 1, 2, \ldots, n$*, and* $n^{-1} \sum_{i=1}^n v_i \le v$*. Then, for all* $0 < \varepsilon < 1$,

$$\mathrm{P}\left(\sum_{i=1}^n Y_i \ge y \right) \le \begin{cases} \exp\left\{ -(1-\varepsilon) \frac{y^2}{2nv} \right\} & y \in [0, \frac{\alpha_\varepsilon nv}{b}] \\ \exp\left\{ -\frac{\alpha_\varepsilon}{b} \left(y - (1+\varepsilon) \frac{\alpha_\varepsilon nv}{2b} \right) \right\} & y \in [\frac{\alpha_\varepsilon nv}{b}, nb] \\ 0 & y \in [nb, \infty) \end{cases}$$

Proof Let $\Psi_i(\cdot)$ be the log of the moment generating function of the conditional distribution (given $\mathcal{F}_i$) of Y_i. Then, for all $t > 0$,

$$\Psi_i(t) = \mathrm{E}(Y_i \,|\, \mathcal{F}_{i-1}) + \tfrac{1}{2} \operatorname{Var}(Y_i \,|\, \mathcal{F}_{i-1}) t^2 + \frac{1}{6} \Psi^{(3)}(\lambda_t t) t^3 \tag{2.1}$$

for some $0 \le \lambda_t \le 1$. But, since $\mathrm{E}\, e^{\lambda_t t Y_i} \ge 1$,

$$\begin{aligned} |\Psi^{(3)}(\lambda t)| &\le \frac{\mathrm{E}(Y_i^3 e^{\lambda t Y_i} \,|\, \mathcal{F}_{i-1})}{\mathrm{E}(e^{\lambda t Y_i} \,|\, \mathcal{F}_{i-1})} \\ &\le e^{tb}\, \mathrm{E}(|Y_i|^3 \,|\, \mathcal{F}_{i-1}) \\ &\le b v_i e^{tb}. \end{aligned} \tag{2.2}$$

Conditioning on $\mathcal{F}_{i-1}$ clearly plays no role here so combining (2.1) and (2.2) we obtain part i). To prove part ii) note that we have,

$$\Psi_i(t) \leq \bar{\Psi}_i(t) \equiv \frac{1}{2} v_i t^2 + \frac{1}{6} b v_i e^{tb}.$$

Hence, for any $t > 0$,

$$\begin{aligned}
\mathrm{P}\left(\sum_{i=1}^n Y_i > y\right) &\leq e^{-ty}\,\mathrm{E}\left(e^{t\sum_{i=1}^n Y_i}\right) \\
&= e^{-ty}\,\mathrm{E}\left(e^{t\sum_{i=1}^{n-1} Y_i}\,\mathrm{E}\left(e^{tY_n} \mid \mathcal{F}_{n-1}\right)\right) \\
&\leq e^{-ty+\bar{\Psi}_n(t)}\,\mathrm{E}\left(e^{t\sum_{i=1}^{n-1} Y_i}\right).
\end{aligned}$$

Continue by induction to obtain.

$$\begin{aligned}
\mathrm{P}\left(\sum_{i=1}^n Y_i > y\right) &\leq e^{-ty} e^{\sum_{i=1}^n \Psi_i(t)} \\
&\leq e^{-ty+mvt^2/2+mbve^{bt}t^3/6}.
\end{aligned}$$

Now, if $0 \leq y \leq \alpha_\varepsilon nv/b$ take $t = y/(nv)$ and note

$$\begin{aligned}
\frac{y^3 b}{6n^2 r^2} e^{by/nv} &\leq \frac{y^2}{6nv} \alpha_\varepsilon e^{\alpha_\varepsilon} \\
&= \frac{\varepsilon Y^2}{2nv}.
\end{aligned}$$

Therefore, in this range,

$$\begin{aligned}
\log \mathrm{P}\left(\sum_{i=1}^n Y_i \geq y\right) &\leq -\frac{y^2}{2nv} + \frac{y^3 b}{6n^2 v^2} e^{yb/nv} \\
&\leq -\frac{1}{2}(1-\varepsilon)\frac{y^2}{nv}.
\end{aligned}$$

To obtain the result for the range $\alpha_\varepsilon nv/b < y \leq nb$ take $t = \alpha_\varepsilon/b$.

□

The proof of the theorem also uses the fact that $\tilde{W}_1, \tilde{W}_2, \ldots, \tilde{W}_n$ is a martingale difference sequence. There are, however, two main differences between the two proofs. The first is that $\tilde{W}_i$ is trivially bounded only by $\mathrm{O}(i)$ which is too large be useful. But given X_i, $\tilde{W}_i$ itself is a sum of $i-1$ iid random variables and hence is actually of order $\sqrt{i}$. We will use lemma 2.1 to claim that with high enough probability $\tilde{W}_i = \mathrm{O}\left(\sqrt{i}\right)$ uniformly in i. Secondly, the proof of the lemma was quite simple since the conditional variance of Y_i is non-stochastic. This is not true for the $\tilde{W}_i$ sequence:

$$\begin{aligned}\mathrm{Var}(\tilde{W}_i \,|\, \mathcal{F}_{i-1}) &= \mathrm{Var}\left(\sum_{j=1}^{i-1} h(X_i, X_j) \,|\, \mathcal{F}_{i-1}\right) \\ &= \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \int_0^1 h(x, X_j) h(x, X_k)\, dx,\end{aligned}$$

which is itself a U-statistic. This means that, in the proof, after taking care of the i-th term, we have to consider the characteristic function of a *new* U- statistic defined similarly but with a different kernel which is a function of $X_1, X_2, \ldots, X_{i-1}$ only. Here is the formal proof.

Proof of Theorem 1.1 Consider the analogue to step (2.1) of Lemma 2.1. By (2.2)

$$\begin{aligned}&\mathrm{E}\left(e^{t\sum_{j=1}^{n-1} h(X_n, X_j)} \,|\, \mathcal{F}_{n-1}\right) \\ &\quad\leq \exp\left\{\frac{1}{2}t^2 \sum_{j=1}^{n-1}\sum_{k=1}^{n-1} \int_0^1 h(x, X_j) h(x, X_k)\, dx + a_n\right\},\end{aligned}$$

where a_n is some bound derived from the bound on the sum. Hence

$$
\begin{aligned}
&\mathrm{E}\left(e^{t\sum_{i=2}^{n}\sum_{j=1}^{i-1}h(X_i,X_j)}\right)\\
&\quad\le\ \mathrm{E}\left(e^{t\sum_{i=2}^{n-1}\sum_{j=1}^{n-1}\left(h(X_i,X_j)+\int h(x,X_i)h(x,X_j)\,dx\right)+\frac{t^2}{2}\sum_{i=1}^{n-1}h^2(x,X_i)\,dx+a_n}\right).
\end{aligned}
$$

The first step in the proof is to define these new kernels that appear in the induction step and establish some of their properties. Let $g_0(x,y) = \beta e^{-\beta\rho}(n\|h\|_*)^{-1}h(x,y)$ for some $0 < \beta \le \rho^{-1}$ be a normalized version of the original kernel. Let $f_0(\cdot\,) = 0$ and define the functions $\bar{g}_i(\cdot\,,\cdot\,)$, $g_i(\cdot\,,\cdot\,)$, and $f_i(\cdot\,)$, $i = 1,2,\ldots,n$, recursively as follows.

$$
\begin{aligned}
\bar{g}_i(x,y) &\equiv \mathrm{E}\left(g_i(x,X)g_i(y,X)\right),\\
g_{i+1}(\cdot\,,\cdot\,) &\equiv g_i(\cdot\,,\cdot\,)+\bar{g}_i(\cdot\,,\cdot\,)\\
f_{i+1}(\cdot) &\equiv f_i(\cdot)+\mathrm{E}(g_i(\cdot\,,X)f_i(X))+\tfrac{1}{2}\,\mathrm{E}(g_i^2(\cdot\,,X)). \qquad (2.3)
\end{aligned}
$$

Note that for all $i = 0,1,\ldots,n$, $g_i(\cdot\,,\cdot\,)$ is a symmetric kernel and

$$\mathrm{E}\,g_i(X,\cdot) = 0. \qquad (2.4)$$

We are now going to bound these functions. Let $t = \beta e^{-\beta\rho}/n$. Since

$$
\begin{aligned}
\bar{g}_0(x,y) &= t^2\int_0^1\left(\sum_{m=1}^{\infty}\nu_m\phi_m(x)\phi_m(t)\right)\left(\sum_{m=1}^{\infty}\nu_m\phi_m(y)\phi_m(t)\right)dt\\
&= t^2\sum_{m=1}^{\infty}\nu_m^2\phi_m(x)\phi_m(y),
\end{aligned}
$$

we obtain that $g_1(x,y) = \sum_{m=1}^{\infty}(t\nu_m + t^2\nu_m^2)\phi_m(x)\phi_m(y)$. A recursive argument yields

$$g_i(x,y) = \sum_{m=1}^{\infty} \nu_{i,m}\phi_m(x)\phi_m(y), \qquad \mathrm{i} = 0, 1, \ldots, n,$$

where $\nu_{0,m} = \beta e^{-\beta\rho} n^{-1}\nu_m$, $m = 1, 2, \ldots$ and

$$\nu_{i+1,m} = \nu_{i,m} + \nu_{i,m}^2, \qquad \mathrm{i} = 0, 1, \ldots, n-1,\ m = 1, 2, \ldots.$$

We prove now that

$$|\nu_{i,m}| \le |\nu_{0,m}| e^{\beta\rho i/n}, \qquad \mathrm{i} = 0, 1, \ldots, n,\ m = 1, 2, \ldots. \tag{2.5}$$

That (2.5) holds for $i = 0$ is trivial. We proceed to show, by induction, that it holds for $1 \le i \le n$. Suppose that (2.5) holds for some i, $0 \le i < n$, then for any m

$$\begin{aligned} |\nu_{i+1,m}| &\le |\nu_{i,m}| + \nu_{i,m}^2 \\ &\le |\nu_{0,m}| e^{\beta\rho i/n}(1 + |\nu_{0,m}| e^{\beta\rho}). \end{aligned}$$

But,

$$|\nu_{0,m}| = t|\nu_m| \le \beta e^{-\beta\rho} n^{-1}\rho. \tag{2.6}$$

Hence

$$\begin{aligned} |\nu_{i+1,m}| &\le |\nu_{0,m}| e^{\beta\rho i/n}(1 + \beta\rho i/n) \\ &\le |\nu_{0,m}| e^{\beta\rho(i+1)/n}. \end{aligned}$$

Equation (2.5) follows. Now, (2.5) implies that

$$\|g_i\|_*^2 = \sup_x \int_0^1 g_i^2(x,y)\, dy$$

$$
\begin{aligned}
&= \sup_x \sum_{m=1}^{\infty} \nu_{i,m}^2 \phi_i^2(x) \\
&\leq e^{2\beta\rho i/n} \sup_x \sum_{m=1}^{\infty} \nu_{0,m}^2 \phi_m^2(x) \\
&= e^{2\beta\rho i/n} \|g_0\|_*,
\end{aligned}
$$

or,

$$
\|g_i\|_* \leq \frac{\beta}{n} e^{-\beta\rho(n-i)/n}, \qquad \mathrm{i} = 0, 1, \ldots, n-1. \tag{2.7}
$$

Moreover,

$$
\begin{aligned}
\|g_{i+1}\|_\infty &\leq \|g_i\|_\infty + \|\bar{g}_i\|_\infty \\
&\leq \|g_i\|_\infty + \|g_i\|_*^2 \\
&\leq \|g_0\|_\infty + \frac{\beta^2}{n^2} \sum_{j=0}^{i-1} e^{-2\beta\rho(n-i)/n}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\|g_i\|_\infty &\leq \frac{\beta e^{-\beta\rho} \|h\|_\infty}{n\|h\|_*} + \frac{\beta^2 e^{-2\beta\rho} (e^{2\beta\rho i/n} - 1}{n^2 (e^{2\beta\rho/n} - 1)} \qquad (2.8) \\
&\leq \frac{\beta e^{-\beta\rho} \|h\|_\infty}{n\|h\|_*} + \frac{\beta^2}{n}, \qquad \mathrm{i} = 0, 1, \ldots, n-1.
\end{aligned}
$$

It follows from (2.3) that $\mathrm{E}\, f_i(X)$ is an increasing sequence and

$$
\mathrm{E}(f_{i-1}(X)) \leq \mathrm{E}(f_i(X)) + \tfrac{1}{2} \|g_i\|_*,
$$

and hence $|\,\mathrm{E}(f_{i-1}(X))| \leq |\,\mathrm{E}(f_i(X))| + \frac{1}{2}\|g_i\|_*$. An argument similar to (2.8) yields

$$
|\,\mathrm{E}(f(X_i))| \leq \frac{\beta^2}{2n}. \tag{2.9}
$$

Next we bound the L_2 norm of f_i. For that write $f_i = \sum_{m=1}^{\infty} \zeta_{i,m}\phi_m$ and $\int_0^1 g_i^2(x,y)\,dy = \sum_{m=1}^{\infty} \omega_m \phi_m(x)$. Note that

$$\left(\sum_{m=1}^{\infty} \omega_m^2\right)^{\frac{1}{2}} \leq \|g_i\|_*^2 \tag{2.10}$$

Now, multiply both sides of (2.3) by ϕ_m and integrate to obtain

$$\zeta_{i+1,m} = \zeta_{i,m} + \nu_{i,m}\zeta_{i,m} + \tfrac{1}{2}\omega_m, \qquad \text{i} = 0,1,\ldots,n-1,\ m = 1,2,\ldots,$$

so

$$|\zeta_{i+1,m}| \leq (1 + |\nu_{m,i}|)|\zeta_{i,m}| + \tfrac{1}{2}|\omega_m|. \tag{2.11}$$

It follows from (2.5), (2.6), (2.7), (2.10), and (2.11) that

$$\begin{aligned}
\|f_{i+1}\|_2 &\leq \left(1 + \frac{\beta\rho}{n}\right)\|f_i\|_2 + \frac{\beta^2}{2n^2} \\
&\leq \frac{\beta^2}{2n^2}\sum_{j=0}^{i-1}\left(1 + \frac{\beta\rho}{n}\right)^j \\
&\leq \frac{\beta^2(e^{\beta\rho i/n} - 1)}{2n\beta\rho}.
\end{aligned}$$

Finally bound $\|f_i\|_\infty$. We use the above bound on the L_2 norm together with (2.7) to obtain:

$$\begin{aligned}
\|f_{i+1}\|_\infty &\leq \|f_i\|_\infty + \tfrac{1}{2}\|f_i\|_2\|g_i\|_* + \|g_i\|_*^2 \qquad (2.12)\\
&\leq \|f_i\|_\infty + \frac{\beta^2(e^{\beta\rho} - 1)}{2n\beta\rho}\frac{\beta}{n}e^{-\beta\rho(n-i)/n} + \frac{\beta^2}{2n^2}e^{-2\beta\rho(n-i)/n} \\
&\leq \frac{\beta^3(e^{\beta\rho i/n} - 1)}{2n^2(e^{\beta\rho/n} - 1)} + \frac{\beta^2 e^{-2\beta\rho}(e^{2\beta\rho i/n} - 1)}{2n^2(e^{2\beta\rho/n} - 1)} \\
&\leq \frac{\beta^2(1 + \beta e^{\beta\rho})}{2n}.
\end{aligned}$$

Let $W_i = \sum_{j=1}^{i-1} g_{n-i}(X_i, X_j)$, $i = 1, \ldots, n$. These random variables are the "modified" U-statistics which were mentioned in the introduction to the main body of the proof. We give a uniform bound on their values. We obtain from Lemma 2.1, (2.4), (2.7), and (2.8) that

$$\mathrm{P}(|W_i| > d_n\beta/\sqrt{n}) \le 2e^{\frac{1}{2}(1-\varepsilon)d_n^2}. \tag{2.13}$$

It follows from Markov's inequality that

$$\mathrm{P}\left\{\mathrm{P}(|W_i| > d_n\beta/\sqrt{n} \,|\, \mathcal{F}_{i-1}) \le 2e^{-\frac{1}{4}(1-\varepsilon)d_n^2}\right\} \le e^{-\frac{1}{4}(1-\varepsilon)d_n^2}. \tag{2.14}$$

Define now

$$\tilde{W}_i = \begin{cases} W_i & \begin{array}{l} |W_i| \le d_n\beta/\sqrt{n}, \qquad \text{and} \\ \mathrm{P}(|W_i| \ge d_n\beta/\sqrt{n} \,|\, \mathcal{F}_{i-1}) > \exp\{-\frac{1}{4}(1-\varepsilon)d_n^2\} \end{array} \\ 0 & \text{otherwise} \end{cases} .$$

Let A_i be the indicator of the event $\{\tilde{W}_j = W_j : j \le i\}$. We obtain from (2.13) and (2.14) that

$$\mathrm{P}\left(\sum_{i=2}^{n} A_i \sum_{j=1}^{i-1} h(X_i, X_j) \ne \sum_{i=2}^{n}\sum_{j=1}^{i-1} h(X_i, X_j)\right) \le 3ne^{-\frac{1}{4}(1-\varepsilon)d_n^2}. \tag{2.15}$$

Now, since by (2.4) and (2.7) $\mathrm{E}(W_i) = 0$ and $|W_i| \le i\|g_i\|_\infty \le iC_1/n$, we obtain that

$$|\,\mathrm{E}(\tilde{W}_i \,|\, \mathcal{F}_{i-1})| \le C_1 i n^{-1} \exp\{-\tfrac{1}{4}(1-\varepsilon)d_n^2\}, \tag{2.16}$$

and by (2.12) and (2.15)

$$\frac{\mathrm{E}(|f_{n-i}(X_i) + \tilde{W}_i|^3 e^{f_{n-i}(X_i)+\tilde{W}_i} \,|\, \mathcal{F}_{i-1})}{\mathrm{E}(e^{f_{n-i}(X_i)+\tilde{W}_i} \,|\, \mathcal{F}_{i-1})} \le \left(\frac{\beta^2(1+\beta e)}{2n} + \frac{d_n\beta}{\sqrt{n}}\right)^3.$$

Since $A_{i-1}=0$ implies that $A_i=0$,

$$\begin{aligned}
&\mathrm{Var}(A_i(f_{n-i}(X_i)+W_i^2)\,|\,\mathcal{F}_{i-1}) \qquad (2.17)\\
\le\;& A_{i-1}\,\mathrm{E}\left((f_{n-i}(X_i)+W_i)^2\,|\,\mathcal{F}_{i-1}\right)\\
=\;& A_{i-1}\Bigg(\|f_{n-i}\|_2^2+2\sum_{j=1}^{i-1}\int_0^1 f_{n-i}(x)g_{n-i}(x,X_j)\,dx\\
&+2\sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\bar{g}_i(X_j,X_k)\,dx+\sum_{j=1}^{i-1}\int_0^1 g_i^2(X_j,x)\,dx\Bigg).
\end{aligned}$$

We obtain from 2.1, (2.9), (2.12), and (2.16)–(2.17) that

$$\begin{aligned}
&\mathrm{E}\left(e^{A_i f_{n-i}(X_i)+\tilde{W}_i}\,|\,\mathcal{F}_{i-1}\right)\\
\le\;& \exp\Bigg\{\frac{\beta^2}{2n}+\frac{C_1 i e^{-\frac{1}{4}(1-\varepsilon)d_n^2}}{n}+\frac{1}{6}\left(\frac{\beta^2(1+\beta e)}{2n}+\frac{d_n\beta}{\sqrt{n}}\right)^3\\
&+\frac{1}{2}A_{i-1}\Bigg(\|f_{n-i}\|_2+2\sum_{j=1}^{i-1}\int_0^1 f_{n-i}(x)g_{n-i}(x,X_j)\,dx\\
&+2\sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\bar{g}_i(X_j,X_k)\,dx+\sum_{j=1}^{i-1}\int_0^1 g_i^2(X_j,x)\,dx\Bigg)\Bigg\}.
\end{aligned}$$

Recall that $A_1\ge A_2\ge\cdots\ge A_n$. We obtain

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\sum_{j=2}^{i}A_j\left(f_{n-i}(X_j)+\sum_{k=1}^{j-1}g_{n-i}(X_j,X_k)\right)\right\}\,|\,\mathcal{F}_{i-1}\right] \qquad (2.18)\\
\le\;& \exp\Bigg\{\frac{\beta^2}{2n}\frac{C_1 i e^{-\frac{1}{4}(1-\varepsilon)d_n^2}}{n}+\left(\frac{\beta^2(1+\beta e)}{2n}+\frac{d_n\beta}{\sqrt{n}}\right)^3\\
&+\sum_{j=2}^{i-1}A_j\left(f_{n-i+1}(X_j)+\sum_{k=1}^{j-1}g_{n-i+1}(X_j,X_k)\right)\Bigg\}
\end{aligned}$$

Use (2.18) beginning with $i=n$ and go back to obtain that

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\sum_{i=2}^{n}A_i\sum_{j=1}^{i-1}g_0(X_i,X_j)\right\}\right]\\
\le\;& \exp\left\{\tfrac{1}{2}\beta^2+\tfrac{1}{2}C_1(n-1)e^{-\frac{1}{4}(1-\varepsilon)d_n^2}+\frac{1}{6}n^{-\frac{1}{2}}\left(\frac{\beta^2(1+\beta e)}{2\sqrt{n}}+d_n\beta\right)^3\right\}.
\end{aligned}$$

Recall that $g_0 = \beta e^{-\beta\rho}\|h\|_*^{-1}h$ and use Markov's inequality to obtain that

$$\begin{aligned}
&\mathrm{P}\left(n^{-1}\sum_{i=2}^{n} A_i \sum_{j=1}^{i} h(X_i, X_j) > y\right) \qquad (2.19)\\
&\leq \quad \exp\left\{-\frac{\beta e^{-\beta\rho}}{\|h\|_*}y + \tfrac{1}{2}\beta^2 + \tfrac{1}{2}C_1(n-1)e^{-\frac{1}{4}(1-\varepsilon)d_n^2} + n^{-\frac{1}{2}}\left(\frac{\beta^2(1+\beta e)}{2\sqrt{n}} + d_n\beta\right)^3\right\}
\end{aligned}$$

The theorem follows from (2.15) and (2.19).

□

3 Application for testing.

We apply the main result, Theorem 1.1 to a family of test statistics that are useful for testing goodness of fit to the uniform distribution. We descrie this application in detail in Bickel and Ritov (1992).

Let $h_\omega(\cdot\,,\cdot)$, $\omega \in \Omega$ be a family of kernels satisfying the following assumptions:

(K1) $h_\omega(x, y) \equiv h_\omega(y, x)$ and $\int_0^1 h_\omega(\cdot\,, y)\,dy = 0$.

(K2) $\|h_\omega\|_\infty = \mathrm{O}\,(w)$, $\|h_\omega\|_* = \mathbf{\Omega}(\sqrt{w})$, and $\rho(h_\omega) = \mathrm{O}\,(\sqrt{w})$. where $a_n = \mathbf{\Omega}(b_n)$ denotes that $a_n = \mathrm{O}\,(b_n)$ and $b_n = \mathrm{O}\,(a_n)$.

(K3) $\Omega = \{1, 2, \ldots\}$ or $\Omega = [\omega_0, \infty)$. In the latter case, $\|\omega_1^{-1}h_{\omega_1} - \omega_2^{-1}h_{\omega_2}\|_\infty \leq c_1|\omega_1 - \omega_2|/\omega_1$, $\|\omega_1^{-1}h_{\omega_1} - \omega_2^{-1}h_{\omega_2}\|_* \leq c_2|\omega_1 - \omega_2|/\omega_1^{3/2}$, $\|\omega_1^{-1}h_{\omega_1} - \omega_2^{-1}h_{\omega_2}\|_* \geq c_3|\omega_1 - \omega_2|/\omega_1^{3/2}$, and $\rho(\omega_1^{-1}h_{\omega_1} -$

$\omega_2^{-1}h_{\omega_2}) \le c_4|\omega_1 - \omega_2|/\omega_1^2$, for all $\omega_2 > \omega_1 > \omega_0$ and $\omega_2 - \omega_1 < 1$ and some positive constants $c_1, \dots c_4$.

We consider the following family of statistics:

$$T_\omega = \frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1} h_\omega(X_i, X_j), \qquad \omega \in \Omega.$$

(T_ω depends, of course, explicitly on n.)

Such a class of test statistics can be derived using a maximum likelihood idea. We can consider $\mathbb{F}$ the family of all continues alternatives to the uniform distribution as a parametric sieve of submodels. That is, $F_0 \subset \mathbb{F}_1 \subset \cdots \subset \mathbb{F}$ where F_0 is the uniform distribution and $\mathbb{F}_j$ are regular j dimensional parametric sub-models and $\overline{\bigcup_j \mathbb{F}_j} = \mathbb{F}$ and the closure is take in (say) the Hellinger metric. We can parameterize each $\mathbb{F}_j$ by $\vartheta_{[j]} \equiv (\vartheta_1, \dots \vartheta_j)$ such that if the densities corresponding to $\mathbb{F}_j$ are $\{f(\cdot, \vartheta_{[j]} : \ \vartheta_{[j]} \in R^j\}$ and

$$l_j(X) \equiv \frac{\partial}{\partial \vartheta_j} \log f(X, \vartheta_{[j]})|_{\vartheta_{[j]=0}}$$

then $\{1, l_1, l_2, \dots, \}$ is an orthonormal basis to $L_2[0,1]$. Let

$$T_{jn} = \sum_{m=1}^{j}\left(n^{-1/2}\sum_{i=1}^{n} l_m(X_i)\right)^2 - j.$$

Then the tests which reject for large values of T_{jn} are asymptotically maxmin for testing F_0 vs. $\{F : \ F \in \mathbb{F}_j, \mathcal{H}(F, F_0) \le c\}$ where $\mathcal{H}$ is the Hellinger distance. T_{jn} is the Neyman smooth test for this problem, Neyman (1942). The χ^2 family of tests is an important example.

Mann and Wald (1942) argued for using the standard χ^2 statistics with $k_n = \Omega(n^{1/5})$ but this prescription seems unsatisfactory — see Kallenberg, Oosterhoff, and Schriever (1985). Rayner and Best(1989) considered this type of tests, and propose to reject when $T_{jn} \geq a_{jn}$ for some j and suitable selected sequences $a_{jn} \nearrow \infty$. Bickel and Ritov (1992) considered this family further and proved that it has a weak kind of efficiency. If l_j are uniformly bounded then these statistics satisfy conditions (K1)–(K3) with $h_j(x, y) \equiv \sum_{m=1}^{j} l_m(x)l_m(y)$. In particular, $\|h_j\|_\infty \geq \sum_{m=1}^{j} \|l_m\|_\infty^2$, $\|h_j\|_*^2 = \sup_x \sum_{m=1}^{j} l_m^2(x)$, and $\rho(h_j) = \|h_j\|_*^{-1}$.

We also consider a more general class of test statistics. Let $\tilde{f} = n^{-1}\sum_i K_\omega(x, X_i)$ be an estimator of the density. The kernel K_ω satisfies, naturally, $\int_0^1 K(x, \cdot\,)\,dx \equiv \int_0^1 K(\cdot\,, y)\,dy \equiv 1$. Then a possible χ^2-type statistic for testing uniformity is $\int (\tilde{f}(x) - 1)^2\,dx$, which is equivalent to T_ω with

$$h_\omega(x, y) \equiv \int_0^1 K_\omega(z, x)K_\omega(z, y)\,dz - 1.$$

Note that the standard χ^2 statistic which is based on dividing the interval $[0, 1]$ into k subintervals of equal length has this structure with $\omega = k$ and $h_\omega(x, y) = \omega\mathbb{1}([x/\omega] = [y/\omega]) - 1$, where $\mathbb{1}$ is the indicator function and $[x]$ denotes the larger integer not greater than x. In other cases, $K_\omega \sim \omega K(\omega(y - x))$ (with some modification to take the finite support into account) For example we can take to modify the family described above by

$$K_\omega(x,y) = \omega\left(f(w(x-y)) + f(w(x+y)) + f(w(2-x-y))\right),$$

where f is a probability density function with finite support and symmetric about 0. Conditions (K1)– (K3) are natural in this situation. Proposition 1 below is useful for verifying condition(K2). A similar results holds for condition (K3).

Proposition 3.1 *Suppose*

$$\omega \underline{K}(\omega(x-y)) \le K_\omega(x,y) \le \omega \bar{K}(w(x-y))$$

for $x, y \in (0,1)$, and some positive bounded functions $\underline{K}$, $\bar{K}$. Then h_ω satisfies (K2).

Proof First note that $\omega \underline{K}^{*2}(\omega(x-y)) - 1 \le h_\omega(x,y) \le \omega \bar{K}^{*2}(\omega(x-y)) - 1$, where K^{*2} is the convolution of K with itself. Hence $\|h_\omega\|_\infty = \mathrm{O}(\omega)$ and $\|h_\omega\|_* = \mathbf{\Omega}(\sqrt{\omega})$. Next, fix $x_0 \in (0,1)$ and let $a_\omega = \omega^2 \int_0^1 (\underline{K}^{*2}(\omega(x_0 - y))\, dy = \omega(\omega)$. Finally, let $\{(\nu_{\omega m}, \phi_{\omega m}),\ m = 1, 2, \ldots$ be the orthonormal eigen system of $\|h\|_*^{-1} h_\omega$. Extend ϕ_m to the all real line to be 0 outside $[0,1]$. Then

$$\begin{aligned}
\nu_m &= \|h_\omega\|_*^{-1} \int_0^1 \int_0^1 h_\omega(x,y)\phi_m(x)\phi_m(y)\, dx\, dy \\
&\le \|h_\omega\|_*^{-1} \int_{-1}^1 \int_0^1 |h_\omega(x, x+t)|\, |\phi_m(x)|\, |\phi_m(x+t)|\, dx\, dt \\
&\le \|h_\omega\|_*^{-1} \int_{-1}^1 \sup_x h_\omega(x, x+t)\, dt
\end{aligned}$$

$$
\begin{aligned}
&\leq \|h_\omega\|_*^{-1} \int_0^1 (\omega \bar{K}(\omega t) + 1)\, dt \\
&= \mathrm{O}\left(\omega^{-1/2}\right).
\end{aligned}
$$

□

The following theorem establishes the uniformity behavior under H_0 which is needed for the optimality result in Bickel and Ritov (1992).

Theorem 3.1 *Suppose that* h_ω *satisfies conditions (K1)–(K3) and* $X_1, X_2, \ldots, X_n$ *are uniform. Then for any* $\eta \in (0,1)$*:*

$$
\lim_{M\to\infty} \overline{\lim}_{n\to\infty} \mathrm{P}\left(\sup_{\omega_0<\omega<n^{1-\eta}} \frac{T_\omega}{\sqrt{\omega \log \omega}} > M \right) = 0.
$$

Proof We begin with $\omega = \{1, 2, \ldots\}$. Fix any $M > 0$. It follows from (K1)–(K3) that the condition of Corollary 1.1 are satisfied and hence

$$
\begin{aligned}
\mathrm{P}\left(\max_{\omega<n^{1-\eta}} \frac{T_\omega}{\sqrt{\omega \log \omega}} > M \right) &\leq \sum_{\omega=1}^{[n^{1-\eta}]} \mathrm{P}(|T_\omega| > M(\omega \log \omega)^{1/2}) \\
&\leq \sum_{\omega=1}^{[n^{1-\eta}]} \left(a_1 e^{-a_2 M^2 \log \omega} + a_3 n e^{-a_4 n^\eta} \right) \\
&\to 0,
\end{aligned}
$$

as $n, M \to \infty$, where $a_1, \ldots a_4$ are some positive finite constants. The theorem follows. Consider now the case of Ω an interval. Use the previous argument to bound

$$\max_{\omega=1,2,\dots,\omega<\frac{c_1 n}{\log n}} \frac{T_\omega}{(w\log(w))^{1/2}}.$$

Now

$$P\left(\max_{\omega\in(k,k+1)} \frac{T_\omega}{(\omega\log\omega)^{1/2}} > M\right) \quad\leq\quad P\left(\max_{\omega\in(k,k+1)} \omega^{-1}T_\omega > M\sqrt{\frac{\log k}{k}}\right)$$

Consider now $\max_k \max_{t\in(0,1)} |(k+t)^{-1}T_{k+t} - k^{-1}T_k|$. For any $\omega_1, \omega_2 \in (k, k+1)$, $|\omega_2 - \omega_1| = 4^{-m}$, we obtain from corollary 1.1 and condition (K3) that

$$P\left(|\omega_2^{-1}T_{\omega_2} - \omega_1^{-1}T_{\omega_2}| > M2^{-m}\sqrt{\frac{\log k}{k}}\right) \quad\leq\quad e^{-a_5 M 2^m k\sqrt{\log k}} \tag{3.1}$$

Use now (3.1) and a chaining argument to verify that :

$$\begin{aligned} & P\left(\max_\omega \frac{T_\omega}{(\omega\log\omega)^{1/2}} > 2M\right) \\ & \leq \quad P\left(\max_k \frac{T_k}{(k\log k)^{1/2}} > M\right) \\ & \qquad + \sum_k\sum_m 4^m \max_{\omega\in[k,k+1)} P\left(\left|\frac{T_{\omega+4^{-m}}}{\omega+4^{-m}} - \frac{T_\omega}{\omega}\right| > M2^{-m}\sqrt{\frac{\log k}{k}}\right) \\ & \to \quad 0 \end{aligned}$$

as $M\to\infty$.

□

REFERENCES

Bickel, P.J. and Ritov Y. (1992). Testing for goodness of fit: A new approach. In *Nonparametric Stat. and Related Topics* .(E. Saleh Ed.), 51–57. Elsevier.

De La Pena, V. H. (1992) Decoupling and Khintchines inequalities for U- statistics. *Ann. Prob. 21*, 1877–1892.

Eichelsbacher, P. and Löwe (1993): Large deviation principle for m-variate von-Mises-statistics and U-statistics. Unpublished.

Kallenberg, W.C.M., Oosterhoff, J., and Schriever, B.F. (1985). The number of classes in χ^2. *J. Amer. Statist. Assoc. 80*, 959– 968.

Mann, H.B. and Wald, A. (1942). On the choice of the number of intervals in the application of the chi-square test. *Ann. Math. Statist. 13*, 306–317.

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. *The Herlad Cramér Volume*, (Grenander, Ed.), 213–234. Almquist and Wicksell.

Nolan D. and Pollard D. (1988). Functional limit theorem for U-processes. *Ann. Prob. 16*, 1291–1298.

Nolan D. and Pollard D. (1987). U-processes: Rates of convergence. *Ann. Statist. 15*, 788–799.

Rayner, J.C.W. and Best, D.J. (1989). *Smooth tests of goodness of fit*. Oxford Univ. Press.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* J. Wiley, New York.